
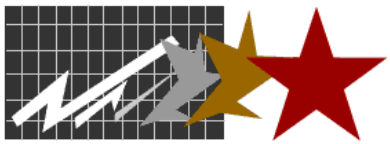

Dillard Research Associates

# Alaska Alternate Assessment

# 2010 - 2011

# Technical Report

Dillard Research Associates
June 30, 2011

# Alaska Alternate Assessment Technical Report

## Table of Contents

# APPENDICES

# EXECUTIVE SUMMARY

The purpose of this Technical Report is to record the administration and reporting of the 2010-2011 Alaska Alternate Assessment.

The critical elements highlighted in that document, with examples of acceptable evidence, include (a) academic content standards, (b) academic achievement standards, (c) a statewide assessment system, (d) validity, (e) reliability, and (f) other dimensions of technical quality. We address the critical elements of training, administration, scoring, and reporting related to the Alaska Alternate Assessment. In addressing technical documentation, we first present content evidence, then reliability, then descriptive statistics and AYP calculations.

In the end, both procedural and empirical evidence are brought to bear for supporting the claim that students with significant cognitive disabilities are assessed in a standardized system of reliable scoring and achieving at various levels of proficiency on the alternate assessment.

Chapters 7, 8 and 9 share the same set of appendices (Appendix 7.1 – 7.6).

---

In each appendix (7.1 through 7.6) are presented in the following order:

7.1 Reading
7.2 Writing
7.3 Mathematics
7.4 Science
7.5 ELOS
7.6 ELOS Frequencies

In the first four appendices in Chapter 7, statistical analyses are presented in this order:

**AYP**

- Participation descriptive statistics at the total test level for each grade or grade band
- Score descriptives for total tests for each subject, including the frequencies of each score (some tables were overly lengthy and were thus not included).

**Test Strand Descriptive Statistics**

- Test Strand descriptives for each subject in the lowest grade or grade band (i.e, grade 3 for grade band 3/4).
- Strand descriptives for each subject in the next lowest grade or grade band (i.e, grade 5 for grade band 5/6).

**Task Item Descriptive Statistcs**

- Operational task descriptives for each subject in the lowest grade or grade band (i.e, grade 3 for grade band 3/4).
- Task descriptives (includes operational and field test items) for each subject in the lowest grade or grade band (i.e, grade 3 for grade band 3/4).

- Task item descriptives for each subject in the next lowest grade or grade band (i.e, grade 5 for grade band 5/6).

**Reliability**

- Item reliability for each subject in each grade band (i.e. grade 3/4).

**ELOS Statistics** (the same ELOS administration protocols applied to all grade bands)

- Descriptive statistics by task and item.

**ELOS Frequencies**

- Score frequencies (1 to 4, indicating level of support required by the student) for each task, each item.

This pattern continues until the highest grade or grade band (10 for 9/10). The subjects always appear in the following order: reading, writing, math, and science.

---

In these tables, the initial tables depict percentages of students participating, then the number of students at each score value for the total sum in a subject area. The following codes apply to grade levels and bands:

w=34, x=56, y=78, and z=910

r=reading, w=writing, m=mathematics, and s=science.

As an example, wtotr = Grade band 3-4 total score in reading.

---

Note: When a number follows the letter 't', it is the task number. When a number follows an underscore (_), it refers to the item number.

As an example, w3tot = Grade 3-4 writing task 3 total score and xt1_1 refers to grade 56, task 1, item 1.

**Field Test Item Summary**

The field test items are displayed within the descriptive statistics for each subject area. It is important to display field test items and operational items together for comparison.

Operational items are underlined and field test items are not underlined.

The field test item means and standard deviations are comparable to the operational items indicating that that the field test items are functioning appropriately. For specific item statistics, please refer to the Appendix for the specific subject area.

# CHAPTER 1: BACKGROUND OF THE ALASKA ALTERNATE ASSESSMENT

## Overview

The 2010–2011 Alaska Alternate Assessment represented an equivalent form test to the 2008–2009 Alternate Assessment. Approximately one-half of the items from the 2008–2009 assessment were paired with equivalent items developed to match the content, construct, and point value of the replaced items. This version of the assessment is referred to as "Form B" and the equivalent items are referred to as "Cousin Items." Sufficient cousin items were developed between 2008 and 2010 to permit as many as six versions of the Alaska Alternate Assessment. Within the Form B assessment, items that were carried forward from the 2008–2009 assessment are called "operational items," while matched cousin items are called "field test items." Operational items have at least two years' of student performance data analysis on the items from 2007-2008 and 2008-2009 tests analysis.

### History of Previous Program

In 2005, a Reliability and Validity study was conducted by Dr. Gerald Tindal which concluded that a need for revision to the State of Alaska's Student Portfolio system in order to meet technical quality requirements set by the No Child Left Behind legislation was required. Following the department's Request for Proposals, Dillard Research Associates was awarded a contract to secure a standardized performance-task assessment for students with significant cognitive disabilities. To provide greater reliability in administration and scoring of the assessment, an online administrator-training program was developed. This online training program includes training and proficiency tests for each subject area. Secure tests were developed in accordance with the State of Alaska's Extended Grade Level Expectations (ExGLEs). Proficiency Level Descriptors (PLDs) were created by teams of content experts.

## Current Program Overview

### Reasons for Current Approach

The current Alaska Alternate Assessment system was developed to meet the requirements of the No Child Left Behind Act of 2001; as described by the National Center on Education Outcomes, alternate assessments are "tools used to evaluate the performance of students who are unable to participate in regular state assessments even with accommodations. Alternate assessments provide a mechanism for students with the most significant cognitive disabilities and for other students who may need alternate assessment formats to be included in the accountability system."

The Alaska Alternate Assessments are standardized performance tasks administered and scored by assessors who undergo a multi-step qualification process. Alaska's current system of assessing students with significant cognitive disabilities has been approved by the U.S. ED Title 1 Final Assessment System Peer Review process.

### Roles of Contractor, Department, and Others

The contractor, Dillard Research Associates (DRA), serves the Alaska Department of Education and Early Development (EED) in developing, training, administering, scoring, and data reporting related to the alternate assessment based on alternate achievement standards (AA-AAS) for students with significant cognitive disabilities. These tasks are defined in greater detail in subsequent sections of this Technical Report.

The EED maintains authority to finalize all deliverable documents, training systems, and reports stemming from the Alaska Alternate Assessment system. The contractor works closely and collegially with personnel in EED's assessment office.

## Summary of Current Program

### Description of Program

The Alaska Alternate Assessments are standardized performance tasks administered and scored by assessors who undergo a multi-step qualification process. The Alternate Assessments are administered to students with significant cognitive disabilities in grades 3-10 (grade bands 3-4, 5-6, 7-8, and 9-10) and measure student achievement in relation to the ExGLEs. All students are assessed in reading, writing, and mathematics. Students in grades 4, 8, and 10 are also assessed in science. The Alaska Alternate Assessments focus on basic academic skills comprised of reading, writing, mathematics, and science tasks that are aligned with Alaska's ExGLEs.

The alternate assessments are comprised of the following components of a web-based training system that can be located at the following URL: http://ak.k12test.com

- Video-based training in each task
- Proficiency examinations
- Practice tests
- Secure test materials accessible only to qualified assessors during the test window
- A data entry and reporting portal

### Description of Students Served

The Alaska Alternate Assessments are administered to students with significant cognitive disabilities in grades 3-10 and measure student achievement in relation to the ExGLEs. Each student's Individual Education Program (IEP) team determines which assessment students in Alaska's Statewide Assessment Program will participate in, based upon criteria established by the EED.

The Alaska Alternate Assessments focus on basic academic skills comprised of reading, writing, mathematics, and science tasks that relate to Alaska's ExGLEs. All eligible students are assessed in reading, writing, and mathematics. Students in grades 4, 8, and 10 are also assessed in science.

### Description of How Scores Are Used

Assessors pre-enter their caseload of students into the online system. After administering the assessments one-on-one to a student, assessors enter student scores directly into the online scoring and reporting system. An unofficial student report is immediately generated for the purpose of providing instructional feedback and guidance to IEP teams. Official student reports that have had the demographic information checked for accuracy and have been assigned proficiency levels were made available to districts on May 16, 2011 via the District Test Coordinators at the secure DRA Web Reporting System. These scores form the basis for Alaska's Adequate Yearly Progress (AYP) report for these students.

Any Standards-Based Assessment (SBA) and Alaska Alternate Assessment receiving a valid score in the content areas of reading, writing, math, and/or science will count toward overall participation and proficiency in the specific content area for AYP. Up to 1% of students attaining proficiency on the Alaska Alternate Assessment may count toward AYP proficiency per district.

## Significant Changes Since Previous Technical Report

Improvements were made to several areas of the Alaska Alternate Assessment for the 2010-2011 testing window, including changes to administration procedures, training, and the online training website. The improvements and changes are detailed in Appendix 1.1.

### *Procedural Changes*

- Expanded the test window to ten weeks to encompass the Terra Nova and SBA testing windows
- Removed "Absent" as a reason that a student was not tested
- Changed the "Mark this area complete" checkbox on the training site
- Added "Refresher Proficiency Test" function
- Alaska's Data Interaction for Alaska Student Assessment (DIASA) will include student data from the Alternate Assessments

### *Training and Continuous Improvement*

- Planned reliability and validity studies
- Recruited Alternate Assessment Advisory Committee members

### *Specific ak.k12test.com Website Changes*

- Organized content into sections
- Encouraged Assessors in training (AITs) to complete an area of training and immediately complete the proficiency assessment for that area, before moving on to the next area of training
- Enhanced training in administering the test and moving to Expanded Levels of Support (ELOS)
- Updated the "Reasons not tested" training page
- Enhanced guidance in scoring writing samples

- Updated all subject-area training sections to improve support materials, especially in the scoring writing section
- Added a section to the Data Entry page for each subject area so that assessors entered data regarding the timing/scheduling accommodations employed with each student for each subject area tested. See Chapter 3.

***Proficiency Testing tab***

- Provided returning Qualified Assessors and Qualified Mentors the opportunity to review a more efficient amount of training pages and participate in a more efficient refresher proficiency assessment process
- Limited AITs to two opportunities to pass the assessment before requiring a reset for both proficiency testing and refresher testing

***Materials Tab***

- Reorganized materials into sections
- Limited access to certain materials sections to Qualified Assessor status or higher
- Reorganized Secure Test Materials into a matrix by Subject Area and Grade Band

***Data Entry Tab***

- Collected additional information for each student assessed

***Reports Tab***

- Changed upload date for Individual Student Reports to May 16, 2011
- Made Unofficial Student Reports available to assessors on June 17, 2011

***Help/Settings Tab (previously ACCOUNT tab)***

- Included Help Desk information at the opening of the Help/Settings tab
- Improved the ability for users to update email addresses within the system
- Collected additional user demographic information

*Appendix 1 Web Changes Handout*

## Organization of Technical Report

The 2011 Technical Report is organized around ten broad topics, with detailed appendices referenced where appropriate. The Report serves as a narrative description of the activities and results of the 2010-2011 testing year. The appendices provide all reference materials, including training agendas, guidance documents, and complete statistical analyses on a variety of required reporting topics.

The topics of the Technical Report are:

Background of the Alaska Alternate Assessment
Test Design and Item/Task Development
Test Administration Procedures
Scoring

Standards Validation
Reporting
Test Validity
Descriptive Statistics
Adequate Yearly Progress (AYP)
Recommended Program Improvements

In all sections where subject area results are reported or described, the Technical Report standardizes the reporting order to: Reading, Writing, Mathematics, and Science.

# CHAPTER 2: TEST DESIGN AND ITEM/TASK DEVELOPMENT

## Overview

The National Center on Education Outcomes http://www.cehd.umn.edu/NCEO/TopicAreas/AlternateAssessments/altAssessTopic.htm describes alternate assessments as "tools used to evaluate the performance of students who are unable to participate in regular state assessments even with accommodations. Alternate assessments provide a mechanism for students with the most significant cognitive disabilities and for other students who may need alternate assessment formats to be included in the accountability system."

The need for developing alternate assessments was in line with the requirements of the Goals 2000 and Improving America's Schools Act (IASA), the Individuals with Disabilities Education Act of 1997 (IDEA), and the IDEA reauthorization in 2004, as well as Alaska's Quality Schools Initiative (QSI), which supported high standards, statewide assessments, and improved results for all students. Until mandated by the federal government, most students with significant cognitive disabilities, and other students with disabilities, were excluded from assessment. Alternate assessments are not typical large-scale assessments, nor are they individualized diagnostic tools. However, the goal is to provide information and accountability for the academic performance of all students in a school district.

The original design of Alaska's Alternate Assessment, a student portfolio, was intended to provide an accountability measure that was consistent with state standards, individualized, performance-based, used independent and reliable scoring, and could be integrated with curriculum and the student's IEP. Students were assessed in language arts, mathematics, and skills for a healthy life. The portfolio assessment was very time consuming for teachers, and teachers often felt that the portfolio measured their ability to construct a portfolio rather than what a student was learning. However, many of the purposes of this first alternate assessment were met. Students were included in the state's comprehensive system of student assessment; student IEPs used academic content standards as goals; students were assessed on academic progress; and, students were included in general education classrooms on a more frequent basis.

After conducting a reliability and validity study, Alaska moved to a performance task assessment that focuses on measuring reading, writing, mathematics, and science. The current Alaska Alternate Assessment uses performance tasks to measure what a student knows and can do in those four core subject areas. The state felt that an assessment with performance tasks offered a more standardized assessment with high technical quality (reliability and validity). Generally, surveys of teachers indicate a greater overall satisfaction with the performance task assessment.

### Description of ExGLEs and their relationship to GLEs

In 1993, the EED developed content standards in English, mathematics, science, geography, history, skills for a healthy life, government and citizenship, fine arts, technology, and world languages. The content standards were broad statements of what students should know and be able to do as a result of their public school experience. A revised edition included content standards for employability, library information/literacy, and cultural standards for students. These content standards are discussed in this document as Grade Level Expectations (GLEs).

In 1999, the Alaska State Board of Education adopted extended performance standards for students with significant cognitive disabilities in the content areas of English/language arts, math, and skills for a healthy life. The reason for developing extended performance standards was to allow for variation in the demonstration of skills across ages and abilities. Different content standards were assigned to, and assessed at, different grade levels.

In response to the 2001 No Child Left Behind legislation, a third edition of the Alaska content standards booklet includes expanded performance standards organized by grade band, called ExGLEs, and revised science content standards and science extended performance standards by grade band. A fourth publication included Alaska history standards.

The No Child Left Behind legislation also required that if a state used AA-AAS's for students with significant cognitive disabilities, "the assessment materials should show a clear link to the content standards for the grade in which the student is enrolled although the grade-level content may be reduced in complexity or modified to reflect pre-requisite skills." In response to this section, the Alaska EED began the process of developing ExGLEs and PLDs.

The ExGLEs are an interpretation of the content standards that should be taught and learned within each grade level. The content is reduced in complexity to provide entry points to the GLEs, while still providing challenging academic expectations for students with significant cognitive disabilities.

## Test Specifications and Blueprint

### Description of Test Specifications (DOTS)

Descriptions of Test Specifications for the 2010–2011 Alternate Assessment are Excel spreadsheets that define all aspects of each item used in the all test materials. In addition to items used in the 2010–2011 assessments, information related to all items used in tests beginning with the 2007–2008 test materials are displayed. Information includes the strand name, the number of answer options, maximum score points, item degree of knowledge, whether the item was an operational or field test item, and statistical data for each item (mean, standard deviation), a statistical analysis of the difficulty of the item (the mean points for each item divided by the maximum points available), and the task weight.

The DOTS documents for reading, writing, mathematics and science contain confidential secure test information and are not available to the public.

### Process of Establishing Test Specifications

The test specifications included the following variables as we developed items:

*Grade Level* – All items were written to appropriate grade bands: 3-4, 5-6, 7-8, 9-10.

*Subject* – All items were written within specific subject area domains: Reading, Writing, Mathematics, and Science.

*Strand Name*: All items were written to fit within subject domains.

*Extended Grade Level Expectation*: These expectations within a content area were organized in content strands and used to organize item writing.

*Item Prompt*: Each item included specific wording for the teacher to use in test administration.

*Item Type*: Both selected and constructed-response items were considered with the vast majority of items using selection responses so that students with physical limitations could participate (respond).

*Item Answer*: Each item was constructed with three options if using a selection type response or an area for the student to construct a response.

*Bias / Content Panel Judgment*: Committee members rated each cousin item as Easy (E), Medium (M) or Hard (H) for students taking the Alaska Alternate Assessment.

*Item Depth of Knowledge*:

Level 1 Rote memory, recall, simple procedure, or apply a one-step, well-defined algorithmic procedure (identify, recall, recognize, use, measure).

Level 2 Some mental processing beyond habitual response. Decisions in how to approach a problem (classify, organize, estimate, display data, compare data).

Level 3 Reasoning, planning, using evidence -- complex and abstract (draw conclusions, cite evidence, explain in terms of concepts, decide which concepts to apply to solve a complex problem). More than one answer, and student has to justify their response.

Level 4 Complex reasoning, planning, developing and thinking, most likely over an extended period of time, plus applying significant conceptual understanding and higher-order thinking. Make several connections (relate ideas within the content area or among content areas, and select one approach among many alternatives to

solve the problem). Design and conduct experiments and projects, develop and prove conjectures, make connections, combine and synthesize ideas into new concepts, critique experimental designs.

### Item Content Test Blueprint and Item Specifications

Test construction for the 2011 testing window was designed to closely match the Form A test administered in 2010. The percent of the strands represented in each subject area and at each grade band are displayed in the appendix.

*Appendix 2.1 Strand Representation: Reading, Writing, Mathematics, and Science*

## Linear Equating

The Reading, Writing, and Mathematics Alternate Assessments tests had significant changes from 05-06 to this year, and the scores are calculated between the two to reach established standard setting cut scores. The point totals of the 2005-2006 assessments available within each strand were established as the baseline. Point totals within each strand in the 2009-10 assessments were compared to the baseline, and a weighting factor calculated.

For instance, assume the 2006 strand 1.34 Numeration totaled 30 possible points and the point total in the same strand for the 2010 test totaled 28, the weighting factor would be .93.

Because the science test was first employed in 2008-2009 and is not based on the 2005-2006 assessment (and the number of points available are the same from 08-09 to 09-10), the Science Alternate Assessment didn't need to undergo linear equating.

*Appendix 2.2 Weights: Reading, Writing, and Mathematics*

### Proficiency Level Descriptor Development

Prior to the adoption of the new Alaska Alternate Assessment, the extended performance standards needed to be revised to reflect the change in the general education academic standards. The existing proficiency level descriptors for the Alternate Assessment Portfolio were universal descriptors. The department assembled teams of content and special education experts, as well as other stakeholders, for the purpose of developing Extended Grade Level Expectations (ExGLEs) for the grade bands 3-4, 5-6, 7-8 and 9-10, and grade-banded Proficiency Level Descriptors based on alternate achievement standards (PLDs) for students with significant cognitive disabilities. The Official Individual Student Reports (ISRs) contain the definitions and descriptions for each proficiency level and at each grade level for each subject area Alaska Alternate Assessment.

*Appendix 2.3 Proficiency Level Descriptors*

### Cut Scores

A standard-setting committee determined cut scores for the new alternate assessment and used the PLDs during that process. During standard setting, the PLDs were revised and were formally adopted by the State Board of Education in July 2007 (reading, writing, and mathematics) and in July 2008 (science). To obtain a proficiency level of advanced, proficient, below proficient, or far below proficient in reading, writing, and mathematics on the Alaska Alternate Assessment, a student must obtain a score as set out in the following tables:

| Reading Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|---|---|---|---|---|
| Advanced | 63 or above | 77 or above | 52 or above | 57 or above |
| Proficient | 32-62 | 46-76 | 33-51 | 43-56 |
| Below Proficient | 8-31 | 11-45 | 12-32 | 22-42 |
| Far Below Proficient | 7 or below | 10 or below | 11 or below | 21 or below |

| Writing Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|---|---|---|---|---|
| Advanced | 76 or above | 67 or above | 76 or above | 82 or above |
| Proficient | 38-75 | 33-66 | 41-75 | 47-81 |
| Below Proficient | 7-37 | 10-32 | 16-40 | 24-46 |
| Far Below Proficient | 6 or below | 9 or below | 15 or below | 23 or below |

| Mathematics Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|---|---|---|---|---|
| Advanced | 62 or above | 61 or above | 74 or above | 81 or above |
| Proficient | 33-61 | 25-60 | 52-73 | 63-80 |
| Below Proficient | 6-32 | 8-24 | 22-51 | 24-62 |
| Far Below Proficient | 5 or below | 7 or below | 21 or below | 23 or below |

| Science Proficiency Level | Grade 4 | Grade 8 | Grade 10 |
|---|---|---|---|
| Advanced | 44 or above | 44 or above | 44 or above |
| Proficient | 24 - 43 | 29 - 43 | 26 – 43 |
| Below Proficient | 12 - 23 | 16 - 28 | 18 – 25 |
| Far Below Proficient | 11 or below | 15 or below | 17 or below |

# Item/Task Development

### Item Writing, including Scoring Guides

A robust set of field test items were designed in 2009-2010 and underwent Content and Bias Review; no new items were written for the 2011 test window. The 2010-2011 Alaska Alternate Assessment test documents were comprised of approximately one-half operational items (items that had two or more years of performance data) and one-half field test items (items that were developed in 2009-2010 to closely match existing operational items).

The training for scoring writing samples was enhanced for the 2011 test window. New guidance tools included a revised rules sheet with examples of scoring decisions.

*Appendix 2.4 Writing Training Documents*

### Item Development and Expanded Levels of Support (ELOS)

Between January and July, in preparation for a major revision in 2011-2012 of the ELOS administration tests, items were developed within grade-bands allowing items to be grouped into tasks. Two to eight items were developed for each task with scoring developed for partial or full credit depending on the item. Because the actual items developed are secure test items, the Item Development Map has been edited to remove secure test items. The resulting document is called the ELOS Item Blueprint.

*Appendix 2.5a – 2.5d ELOS Item Blueprint*

Two types of items were created: Standard items and Expanded Level of Support (ELOS) items. The ELOS items were created to ensure participation and allow the assessor to ascertain a student's level of independence. These items also allow maximum participation for students with the most significant cognitive disabilities and provide information for assessors on what level of support is necessary for the student to interact with the assessment materials.

During the testing windows in 2005-2006, and 2006-2007 test items consisted of new and field test items. In 2007-2008, the secure tests were organized into grade bands and items were constructed by closely aligning the items to the ExGLEs. Crosswalks were created to map the 2006-2007 test items to the 2007-2008 test items for all subject areas. The test design and specifications were applied in four areas of (a) reading, (b) writing, (c) mathematics, and (d) science. Each test included both teacher administration and scoring protocols and student materials.

Each test consists of tasks that are comprised of several items. These items closely align to the ExGLEs.

### Reduction in Complexity, Depth, and Breadth

Due to the federal regulations provided in December 2003, steps were taken to increase the cognitive accessibility of items. This was done by analyzing and removing potential barriers for students with significant cognitive disabilities. This process was used in the development of items and for both administration and scoring and student materials. Simplified language was used in all text. Alignment was ensured between teacher-scripted language and student materials. General test layout was considered from the view of readability and legibility. Specific administration directions were limited to a single page of the Scoring Protocol for ease of administration. Student materials were organized for ease of administration into 'cards' that either could be cut out or masked by the administrator. Pictures were constructed using primarily black and white for minimal complexity. All items were reviewed with administration and development steps toward reducing complexity.

Depth-of-knowledge (DOK) was judged in the analysis of the Alaska Alternate Assessment. An alignment study was conducted in 2007 by Karvenon and Almond; the information was used to guide item adaptations for the 2007-2008 secure test items. Categorical concurrence, range of knowledge, and balance of representation were defined originally by Webb, and adapted by Dr. Tindal for use with students with significant cognitive disabilities, and then defined based on operational use within the Alaska Alignment Study.

Items were developed based on a one-to-one correspondence with the ExGLEs. All strands and attributes were equally addressed in accordance to proportion of points for each task. The total points for each test was fixed at 100 points to allow proficiency standards from the first year to be comparable to the second year of testing. Weighting was needed and an algorithm was used to equalize the differential points across strands/attributes.

### Bias and Sensitivity Review

A bias and sensitivity review of the new test items was conducted in November 2007. During this process, reviewers examined the bias of the assessment and if the format would affect student performance. A group of 12 participants from Alaska and two specialists with the deaf and blind community from Oregon were selected to review all items. All reviewers were given examples to focus on during the review and all held qualified assessor certificates and certification in special education. Items were updated based on the results of this review prior to the 2007-2008 testing window.

A second Bias and Content Committee was convened in September 2009 to analyze cousin items, a pool of new, related items to the existing items in the Alaska Alternate Assessment. The results of this analysis are indicated in the DOTS document.

# Test Design and Development

### Representation and Functionality (New)

The 2009-2010 cousin items reflected minimal construct under-representation or construct irrelevant variance (CIV) to ensure functionality.

- Select the most appropriate word with the least number of syllables
- Reduce number of words used in items, directions, and passages
- Use independent clause structure instead of dependent clause structure in passages
- Develop prompts with minimal wording
- Ensure more opportunities for modeling
- Provide examples when possible
- Create clear (not tricky) distractors
- Provide explicit textual information with reduced requirements for extended inference
- Provide rules rather than exceptions
- Use careful sequencing so that potentially similar/confusing information is not presented
- Place items adjacent to similar information
- Provide multiple choice options for items when possible or appropriate for item construction

Rasch equating was used to ensure functionality by calculating fit statistics that reflect the degree to which ability and difficulty are mapping correctly.

### Psychometric Guidelines for Selecting Items/Tasks for Item/Task Bank

We used traditional guidelines for selecting items and tasks that rely on reliability coefficients but also on implementation in the field. As described in other sections of this report, extensive training of new and returning qualified assessors and mentors was conducted before the testing window opened; in addition, web-based training and proficiency assessments were completed with actual practice in the field required.

We calculated both the mean and standard deviation for each item to ensure the item was functional for a wide range of students.

The entire item bank was developed with all items from the original test completed in 2006-2007 and every year after that to identify common items and use them as anchors for calibrating item values using a Rasch Partial Credit Model.

### Item Field Test and Item Data Review

Field test items for the 2010-2011 tests were not included in calculation of student performance level or district AYP results. However, DRA analyzed the field test items separately; those results are reported in Chapter 10.

### Item Bank Summary

All items that have been used in any version of the Alaska Alternate Assessment, beginning with SY 2006, have been included in the 2011 DOTS, due to be released to EED in November 2011. Student performance on each item is recorded for each year the item was in use. The DOTS is the item bank.

In addition, Rasch modeling completed on items was completed in 2011. See the section regarding Rasch Modeling for more information regarding item performance over time.

## Test Construction

### Steps in the Forms Construction Process

The 2008-2009 Alaska Alternate Assessment served as the baseline document for developing two forms; items included in this version are referred to as operational items. Test items in Form A (2009-2010) were developed by identifying the strand, task, and construct for each operational item and locating matching cousin items for each. This system allows DRA to conduct statistical analysis on the operational items, on the field test items, and equivalent test form analysis. Beginning with testing year 2011-2012, all items in the current item bank are considered operational items.

### Construction of the Operational Forms

Approximately one-half of the 2008-2009 test items (operational items) were replaced by matching cousin items (field test items). Operational items are items that have been used (without modification) in the 2007–2008 and 2008–2009 assessments, and thus have two years of statistical data collected for each item. Field test items were carefully created to match the operational items they would replace.

The test documents for 2010-2011 (Form B) represented the unused half of operational items and field test items.

### Psychometric Guidelines for Constructing Forms

Coverage of Strands (and equal weighting) was used to ensure appropriate and consistent representation of items to strands within each subject area and grade level.

Two forms were developed in 2009-2010 with Form A (where we used half of the items from previous years to serve as operational items and developed cousin items for the other half of the test) and 2010-2011 with Form B (where we again used half of the items from previous years to serve as operational items and developed cousin items for the other half of the test). In all tests, the cousin items were embedded in the actual test but only the operational items were used to calculate Adequate Yearly Progress (AYP).

As we have scaled all the items for each successive year into a common scale using a Rasch Partial Credit Model ($\pi$), we can use item functioning to build alternate forms. All tests were equated between successive years with a nonequivalent group with anchor test

(NEAT) design. The RCPM was used for all analyses. Our intent was to produce item calibrations for the 2009-2010 (Form A) and 2010-2011 (Form B) forms that were equated to the baseline year – the first year the test was administered. We used a chained equating method to link items between multiple years. Our method comprised the following steps:

- Calibrate all items in the baseline year with a RPCM model
- Identify common items between baseline and subsequent years
- Calibrate subsequent years while anchoring common item difficulties and step values to the baseline calibrations

When equating item calibrations in 2009-2010 and 2010-2011 to the baseline year scale, we used any and all common items across years as anchor items, anchoring common items between adjacent years and from previous years. This process allowed more items to be anchored, which ultimately should reduce the equating error.

Table 1 displays the baseline year for each subject and grade. Following Table 1 is a description of the common item-anchoring plan for the two different baseline years.

*Table 1 Baseline Equating Year by Subject and Grade*

| Baseline | Reading | Writing | Math | Science |
|---|---|---|---|---|
| Grade 3-4 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 5-6 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 7-8 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 9-10 | 2006-2007 | **2007-2009** | 2006-2007 | 2007-2009 |

***Baseline 2006-2007.*** Items were first calibrated with a RPCM during 2006-2007, the baseline year for determining linear equating in future years. After calibration, common items between 2006-2007 and 2007-2009 were identified (the same test had been used for two successive years). The 2007-2009 items were then calibrated with the common item difficulty and step values anchored to the 2006-2007 calibrations. In 2009-2010 common items were identified between 2006-2007 and 2009-2010, *and* 2007-2009 and 2009-2010. The 2009-2010 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2006-2007 or 2007-2009). In 2010-2011 common items were identified between 2006-2007, 2007-2009 and 2010-2011, and 2009-2010 and 2010-2011. The 2010-2011 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2006-2007, 2007-2009, or 2009-2010).

***Baseline 2007-2009.*** Because science and grade 9-10 writing were not required Alaska assessments prior to 2007-2009, the baseline years for these assessments were 2007-2009. After the initial calibration for the 2007-2009 years, common items were identified between 2007-2009 and 2009-2010. The 2009-2010 items were then calibrated with the common item difficulty and step values anchored to the 2007-2009 calibrations. In 2010-

2011 common items were identified between 2007-2009 and 2009-2010, and 2010-2011. The 2010-2011 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2007-2009 or 2009-2010).

***Essentially equivalent items.*** When the baseline year was 2006-2007, there were occasionally no common items between the 2006-2007 version of the test and the 2007-2008 version of the test. Yet, our intent was to equate items to the baseline year, 2006-2007. If no common items existed between 2006-2007 and 2007-2008, we identified *essentially equivalent* items. For example, in the following two tasks, the items from 2006-2007 and 2010-2011 consisted of copying letter names; we simply associated items with each other without exact matching of each specific letter with itself across the two time periods.

2006-2007

| Item | Letter | Student Response | Points |
|---|---|---|---|
| 1 | g | | /2 |
| 2 | t | | /2 |
| 3 | k | | /2 |
| 4 | a | | /2 |
| 5 | m | | /2 |
| 6 | B | | /2 |
| 7 | C | | /2 |
| 8 | W | | /2 |
| 9 | I | | /2 |
| 10 | Z | | /2 |

2010-2011

1.34A - Copy Letters - Scoring

| Item | Letter | Student Response | Points |
|---|---|---|---|
| 1 | c | | /2 |
| 2 | f | | /2 |
| 3 | h | | /2 |
| 4 | z | | /2 |
| 5 | w | | /2 |
| 6 | L | | /2 |
| 7 | U | | /2 |
| 8 | J | | /2 |
| 9 | Y | | /2 |
| 10 | V | | /2 |
| 11 | G | | /2 |

Within Appendix 2.6, Tables 3-6 present the "equating map" used for reading by grade-band, Tables 7-10 present the same equating map for writing, Tables 11-14 present the map for math, and Tables 15-18 present the map for Science. When viewing the tables, each row represents a unique item. When the item was presented in multiple years, its variable label appears in each year. The column containing the words "recode" indicates the way the scoring was recoded and was left blank if no recoding was deemed necessary. *Essentially equivalent* items are displayed in bold-faced font.

*Appendix 2.6 Equating Map*

## Data Preparation

Five years of Alaska Alternate Assessment data were prepared for an equating analysis (using Rasch analysis methodology) with data prepared in a similar format for all grade-bands and for all content areas. A number of steps were consistently applied across the subject areas and grade levels that involved cleaning up the data file to exclude missing fields (values) for virtually all items for a small group of students, data aggregation to prepare master files for analysis, and item recoding for ensuring similar scales.

***Systematic deletion of non-responders.*** The Alaska Alternate Assessments contain discontinuation rules for students who are not responding to items aligned with the ExGLEs. The discontinuation rules state that if a student scores a zero on three items within

a task, the remainder of the task should be discontinued and the tester should go on to the next task. If the discontinuation rules are exercised for three consecutive tasks, the test as a whole should be discontinued. The student subsequently is administered the appropriate set of ELOS tasks/items. These rules were in place in the beginning of the 2007-2008 school year.

Careful examination of the data indicated that the discontinuation rules were not always followed. This was true for every year the test was administered. For example, a student may have received a zero on all but two items on the test, but still have been administered every test item. Given that a student such as this should have had the discontinuation rules exercised, we chose to remove these students from all analyses. Students were systematically excluded when greater than 90% of their valid responses across the test items were zeros. Retaining these students' data would likely have resulted in artificially difficult item calibrations because, while the student may have received a valid score of zero, he or she should not have been administered the item in the first place. Our systematic elimination of non-responders from the dataset produced a final sample of students for each item more representative of the intended student population. Although the discontinuation rules were not in place during the 2006-2007 year, we followed the same systematic deletion process in an attempt to obtain the most comparable student samples between years possible for equating. Table 2 shows the number of students in the raw data, and the number remaining after the systematic deletion. In the end, the data file was minimally reduced.

***Data aggregation.*** The alternate assessments were identical for all content areas (reading, writing, math, and science) during the 2007-2008 and 2008-2009 school years. All items were presented in the same format, in the same order, and with the same administration procedures. When applying item response models, such as the Rasch model, larger sample sizes are desirable, as the error associated with each item and person included in the analysis are reduced. Given the identical design of the assessments between 2007-2008 and 2008-2009, and the need for large sample sizes, all item response data were combined between these two years, creating a single "2007-2009" data file.

***Item score recoding.*** There were four reasons that items occasionally needed to be recoded when the item included: (a) a large number of potential scoring options, (b) inconsistent scoring between years for common items, (c) inconsistent item scoring, or (d) missing step values.

In some content areas, particularly writing, a large amount of score reporting options were available. For example, one writing item in grade-band 9/10 had possible scores ranging from 0-35. When such a large number of scores were possible, it became difficult to estimate the step calibrations because very few students received each score. When a large number of scores were possible, the item was recoded into a narrower band of values. For example, the aforementioned 35-point item was recoded to a 7-point scale, with the score options categorized into five 6-point interval ranges and one 5-point interval range (i.e., 0 = 0, 1-6 = 1, 7-12 = 2, 13-18 = 3, 19-24 = 4, 25-30 = 5, and 31-35 = 6).

In other instances, the item scoring changed between years, although the item itself did not change. For example, an item may have been scored dichotomously, 0-1, in one year but then given a partial credit rating, 0-1-2, in the next. If the scoring changed on a common item that needed to be anchored, the items needed to be recoded so they would have identical scoring algorithms. For example, if an item was scored 0-1 in 2006-2007, but 0-1-2 in 2007-2009, the difficulty and step values would be anchored only to the 0-1 values for the 2007-2009 analysis. The model would then not expect students to score a 2 on the item, given that the maximum score on the anchored value is a 1. If these steps had not been employed, estimates on all items would have become skewed.

Occasionally, items had inconsistent scoring protocols. For example, most 6-value scale items in the 2007-2009 version of the test were scored 0-1-2-3-4-5, but one was instead scored 0-5-10-15-20-25. The different scoring for the one item resulted in the item being differentially weighted. All estimates in a Rasch model are based on the total sum score of the test items. If one item contributes more to the sum score, then it is weighted more heavily in the calibrations. Further, a partial credit model assumes there are no empty steps (e.g., the partial credit model would assume that 1-4 were possible score values). All items with inconsistent scoring algorithms were recoded to match the rest of the data.

Finally, on occasion there were items with missing step values. That is, while the item was intended to be scored 0-1-2, only the values of 0-2 were present. In these instances, items were recoded to the least condensed scale possible. For example, if an item was intended to be scored 0-1-2-3-4-5, but only had valid responses in the 0-1-2-3-5 categories, the item would be recoded to 0-1-2-3-4. Items were always recoded to have consistent scoring throughout all five years of the data. The recoding schemes of all items are reported in the equating maps (Tables 2-17 in Appendix 2.6).

**Model and Analysis**

The RPCM is outlined by Masters (1982) as

$$P_{ix}(\theta) = P(X_i = x \mid \theta) = \frac{\exp\left[\sum_{s=0}^{x}(\theta - \delta_{is})\right]}{\sum_{k=0}^{M_i}\exp\left[\sum_{s=0}^{x}(\theta - \delta_{is})\right]} \quad x = 0,1,\ldots,M_i$$

where $P_{ix}$ indicates the probability that a person with ability $\theta$ will receive score $x$ on item $i$.

1. The score $x$ is a count of the successfully completed steps of the item, which is restricted by $M_i$, the number of possible scoring options.
2. All steps, $s$, range from 0 to $M_i$. The probability of receiving score $x$ is a function of the student's estimated ability, $\theta$, and the difficulty of the step values, $\delta_{is}$.
3. As with all Rasch models, the student ability estimates are given by the student's total score and all items are assumed to have equal discrimination values.
4. The step difficulties can be separated from the model and estimated independently. When $\theta = \delta_{is}$, $P_{ix} = 0.5$ and the student has an equal probability of scoring in either of the adjacent scoring categories. Similarly, when $\theta > \delta_{is}$, $P_{ix} > 0.5$, and when $\theta < \delta_{is}$, $P_{ix} < 0.5$.
5. A requirement of the RPCM is that the steps be ordered, so that consecutive step completion leads to higher $\theta$ estimates. However, the $\delta_{is}$ estimates (not $P_{ix}$), are generally not interpretable and consecutive ordering is not required (Yen & Fitzpatrick, 2006).

Winsteps version 3.68 was used for all analyses. Winsteps uses a two-stage estimation process. The first stage consists of a preliminary estimation through PROX, which according to the user manual "capitalizes on the similar shapes of the logistic and normal ogives. [The algorithm] models both the persons and the items to be normally distributed" (Linacre, 2011, p. 488). PROX is run to create initial estimates of person and item locations. These estimates are used as the starting point for the second stage of estimation with joint maximum likelihood estimation (JMLE). Strict convergence criteria were used when running all analyses and no maximum iteration level was set. When equating test forms, common item difficulties and step values were anchored between forms. The non-anchored items were then freely estimated *relative to* the anchored values, by using person estimates from the anchored items to calibrate the non-anchored items (DeMars, 2004). The anchored items thus adjust the calibrations of the non-anchored items to reflect the scale of the anchored items.

*Item maps* for locating common items have been organized by grade bands (3-4, 5-6, 7-8, and 9-10).

The actual item analyses have been organized in an Appendix that includes two types of tables for each subject area and grade band, ordered by item and ordered by measure. To find a specific item in any test year (2006-2007, 2007-2009, 2009-2010, and 2010-2011, use the tables 'ordered by item'. To see the rankings of items within a test and for each specific year, refer to the tables 'ordered by measure'. N.B. Measure refers to the difficulty of the item, with negative numbers reflecting easy items and positive numbers reflecting difficult items. The scale itself ranges from approximately -2 to +2 and is centered at 0.

*Appendix 2.7 Item Order*

*Appendix 2.8 Measure Order*

DRA established two or more years' statistical data on operational items, and one year's data for field-test items. This allows a comparison of the performance metric between operational items and their matched cousin item, which were field-tested.

The results of the analysis of field test items is described in the appendix, and more fully summarized in the next section.

In 2009, DRA and EED constructed a plan that would allow a minimum of six versions of the Alaska Alternate Assessment:

1) AK AA Test (2007-2008, 2008-2009)
2) 1/2 (a) AK AA test + 1/2 (a) FT (FORM A, 2009-2010)
3) 1/2 (b) AK AA test + 1/2 (b) FT (FORM B, 2010-1011)
4) 1/2 (a) AK AA test + 1/2 (b) FT
5) 1/2 (b) AK AA test + 1/2 (a) FT
6) 1/2 (a) FT + 1/2 (b) FT

However, because EED joined a GSEG consortium of states working to create alternate assessments, DRA and EED have agreed to maintain the current two forms of the Alternate Assessment (Form A and Form B, numbers 2 and 3 above) for the future, until such time as the GSEG assessment is developed and ready to be deployed.

If the GSEG assessment is not deployed in Alaska, future forms of the alternate assessments may be constructed using the results of a Rasch Analysis of current items. With the results that are detailed in the Technical Report section regarding Rasch analysis, test forms can be constructed that are:

1. Tied to a 100-point scale, eliminating the need to conduct linear equating; and,
2. Ordered easy to hard, either within a strand or throughout the entire test, without regard to strands.

### Internal Review of the Items and Forms

DRA has established quality assurance steps and protocols designed to eliminate errors in content, grammar and formatting and to improve document retrieval by assigning document-naming protocols to all documents. These protocols are described in the appendix.

*Appendix 2.9 Quality Assurance*

### Test Development Timeline

At the conclusion of the first six-year contract with EED on June 30, 2011, DRA has developed and produced two complete forms of the Alaska Alternate Assessment. Both forms have approximately the same number of tasks and items and represent similar content standards and strands. In addition, a new ELOS test is in development; an anticipated completion date is June 28, 2011.

The new six-year contract, representing 2011-2017, has been established as a maintenance contract, with no new test items to be developed.

# CHAPTER 3: TEST ADMINISTRATION PROCEDURES

## Overview

The Alaska Alternate Assessment is administered by trained qualified assessors, following a standardized scoring protocol. The assessment is administered individually to qualifying students and is scored at the time of administration by the assessor.

## Student Population Tested

This test is reserved for students with significant cognitive disabilities. Individualized Educational Program (IEP) teams make a determination whether a student is eligible to take the Alaska Alternate Assessment by following the guidelines in Alaska's Participation Guidelines for Alaska Students in State Assessments, September 2007 edition, located at: http://www.eed.state.ak.us/tls/assessment/participation_guidelines/ParticipationGuidelinesSept2007.pdf

## Accommodations

The Alaska Alternate Assessment allows accommodations to be utilized during test administration. Accommodations for each student are determined by the student's IEP team.

Accommodations fall into the following categories:

- **Timing/Scheduling** (e.g., extended time, frequent breaks, etc.)
- **Setting** (e.g., study carrel, student's home, separate room, etc.)
- **Presentation** (e.g., repeat directions, read aloud, large print, Braille, etc.)
    - Included with Presentation is **Assistive Devices/Supports** (e.g., calculator, amplification equipment, manipulatives, etc.)
- **Response** (e.g., mark answers in book, scribe records response, point, use an assistive device, etc.)

The Participation Guidelines recommends that an accommodation should be used in the classroom for at least three months prior to testing. This timeline is a suggestion. It is important that the student have practice with the accommodation prior to testing; how much practice will differ by student. This amount of time allows the student to become familiar with the accommodation and ensures that the accommodation is appropriate for the student. A participation guideline is available on the EED website at: http://www.eed.state.ak.us/tls/assessment/participation_guidelines/ParticipationGuidelinesSept2007.pdf

## Test Administrators

Only school personnel may administer the Alaska Alternate Assessment. This includes both teachers and paraprofessionals. In order to become a Qualified Assessor (QA), individuals must participate in online training, pass proficiency tests, and administer a practice assessment that is then reviewed by their Qualified Mentor-Trainer (QT). Each QT must go through this same training, as well as additional in-person training provided annually by the EED, in order to serve as a valuable resource to QAs. These individuals have been appointed by the Special Education Director or Superintendent to be the primary point of contact for EED's Alternate Assessment Program Manager.

## Mentor Responsibilities

A district appoints a person to become a QT. A Mentor-in-training first must meet all of the training requirements to become certified as a QA.

A mentor-in-training attends new mentor training as well as the annual mentor training. Both trainings are provided by EED and include:

- Complete all required training
- Receive materials to support training (PPT, handouts, examples of scoring protocols)
- Train a protégé to become a QA by:
  - Providing orientation to assessments and online training program and ongoing support
  - Reviewing and providing feedback to protégé on practice tests after they achieve proficiency on the online training
  - Upgrading protégé status from AIT to QA, after protégé have produced corrected scoring protocols to the qualifying level

After meeting qualifications, QTs become certified and have their status upgraded by EED or DRA.

Ongoing requirements to continue as a QT:

- Hold a QT Certificate, or attend new QT training
- Attend any required refresher trainings
- Refresh proficiency annually to maintain access to online system
- Sign Test Security Agreements annually and give to EED with a copy on file with the District Test Coordinators (DTCs)

Mentors have access to online reports to track their protégés' progress through training, update their status to QA when appropriate, track progress toward entering student demographic information, progress toward completion of assessment administrations, and to track any assessors who have not completed student assessments during the last week(s) of the testing window.

## Materials

All materials used in training are available to QTs for use in their districts, as they train and certify new QAs. Materials are organized into sections on the ak.k12test.com website. Some material is restricted to personnel with QT status and higher, secure test documents are restricted to personnel with QA status or higher. The training pages and support materials for training are available to all registered users.

## Test Administrator Training

Special education teachers who were selected by their districts to serve as QTs the Alaska Alternate Assessment attended a two-day, New Mentor training on September 27 and 28, 2010. After AITs completed all training and proficiency tests successfully, they administered a practice test, which was reviewed by DRA. Once the AIT completed these tasks, his or her account was updated to the status of QA. During training, these participants also scored a protégé's assessment protocols. After passing all these tasks, participants were upgraded to QT status, and were invited to attend the All Mentor Training in October 2010.

The additional responsibilities of a QT necessitate additional training, which was held October 25-26, 2010 in Anchorage. This training provided more in-depth information on the creation of and changes to the 2010-2011 Alaska Alternate Assessments and Secure website, including training tips to the QTs.

### New Mentor Training

The purpose of the Alaska Alternate Assessment Mentor Program is to prepare district level trainers who train district personnel in correct test administration procedures for the Alaska Alternate Assessment. Mentors are available throughout the year to answer questions and assist district personnel. They are the first point of contact in the district for EED's Alternate Assessment Program Manager. Additionally, mentors act as an advisory group for the Alaska Alternate Assessment. Mentors should be certified teachers in the State of Alaska with a special education endorsement and have experience with low-incidence disabilities. The state encourages every district to have at least one QT and one QA.

The bulk of training occurs on the website http://ak.k12test.com. AITs participate in a series of video vignettes designed to familiarize them with both appropriate testing and scoring techniques. These training vignettes familiarize AITs with the wide variety of tasks they will encounter on the Alaska Alternate Assessment, and demonstrate all the nuances needed in a proper administration. Following the training exercises, AITs must pass a series of brief proficiency tests related to the different tasks in each content area, as well as tests on general administration.

### Summary of Dates and Participants

The New Mentor training was conducted in Juneau, AK, on September 27 and 28, 2010. After a brief introduction, instruction was given in obtaining passwords and login identities, and navigating through the Alaska Alternate Assessment training and score entry website. Participants then completed online training and proficiency tests for each of five content areas: test administration, reading, writing, mathematics, and science.

*Appendix 3.1a New Mentor Training Attendees*

*Appendix 3.1b New Mentor Training Agenda*

*Appendix 3.1c New Mentor Training Handouts*

### Scoring Reliability Analysis

The second day of training was devoted to gaining proficiency in administering the test to a "protégé." Participants administered and scored practice tests to the each other. These materials were collected and rated against a master document for accurate administration and scoring. Participants also rated a "protégé's" assessment in reading, writing, mathematics, science and ELOS. The "protégé" assessment that was scored was a fabricated set of assessments; all AITs received the same set of scored assessments. This allows for consistent scoring across assessors, as well as tracking scoring issues across training years. Twelve of the sixteen participants successfully completed this two-day training regimen. Four participants required additional coaching and practice in scoring a protégé's work.

*Appendix 3.2 New Mentor IRR Report*

*Appendix 3.3a-3.3m Practice Tests*

### Annual Mentor Training

Annual Mentor training was held October 25 and 26, 2010 in Anchorage, AK. Participants from the New Mentor Training in September attended as well as veteran QTs.

*Appendix 3.4a Annual Mentor Training Attendees*

On the first day, QTs engaged in a review of the training website and strategies for training protégés on the Alaska Alternate Assessment.

*Appendix 3.4b Annual Mentor Training Agenda*

*Appendix 3.4c Annual Mentor Training Handouts*

*Appendix 3.4d Annual Mentor Training Writing Scoring Handouts*

On the second day, QTs engaged in training around using strategies and technology to enhance access to the general curriculum for students with significant cognitive disabilities, led by Padmaja Sarathy.

### Webinar

On January 26, 2011, DRA and EED hosted a web-based seminar (webinar) for QTs of the updates to the Alaska Alternate Assessment website. The webinar served to update QTs to improvements to the Alaska Alternate Assessment website (led by DRA) and to update them on procedural information related to the Alaska Alternate Assessment system and procedures (led by EED).

*Appendix 3.5 Webinar*

### Online Training

All assessors must complete the online training through the ak.k12test.com site. After completing training, assessors complete proficiency testing. After participating in training through the ak.k12test site on all aspects of administering, scoring, and data entry for the Alaska Alternate Assessment, Assessors-in-Training (AITs) participate in proficiency testing. Each of five training areas are tested with a 20-question multiple-choice test (Administration, Reading, Writing, Mathematics, and Science). AITs are given two opportunities to earn a passing score of 80% or greater. If the AIT is unsuccessful in two attempts, the AIT must contact his or her Qualified Mentor to reset the proficiency tests. The AIT then has another two opportunities to pass the subject area.

*Appendix 3.6 Training Site Table of Contents*

Qualified Mentors are encouraged to analyze the AIT's performance on the proficiency assessment and compare that to other data available through the ak.k12test.com Web report function.

*Appendix 3.7 Website Report Specifications*

### Refresher training and testing

Returning QAs and returning QTs who completed training in 2009-2010 were eligible to participate in a more efficient training and a refresher-proficiency test. After completing the reduced training sections, Returning Qualified Assessors and Qualified Mentor-Trainers completed a 25-question multiple-choice refresher proficiency test. All five areas of training (Administration, Reading, Writing, Mathematics and Science) were assessed, and a score of 80% was required for passing.

After two attempts at passing the refresher test, a Returning Qualified Assessor's or Qualified Mentor-Trainer's test sessions were re-set to the full set of five proficiency tests.

- Returning Qualified Assessors, were instructed to contact their Mentor to reset the refresher tests.

- Returning Qualified Mentor-Trainers, were instructed to contact EED to reset the refresher tests.

*Appendix 3.8 Refresher Training Tasks*

**Security**

Items and test documents are maintained in a secure fashion. Transfer of items or documents containing secure test items or documents containing FERPA-protected student information is made via a secure file transfer site.

During annual training, all participants are required to sign and return a test security agreement. This document reiterates the message from training: test security is of the utmost importance in obtaining valid and reliable scores. As such, QAs must keep all materials in a confidential location, and refrain from discussing specifics of the test with others. Following the close of the test administration window, all testing materials should be shredded. Teachers cannot access the secure test documents until they have passed the training requirements (passing all proficiency tests and, for assessors-in-training, administration and submission of a practice test). After completion of all requirements, they are granted access to the secure test materials.

The Test Security Agreement is available in the appendix.

*Appendix 3.9 Test Security Agreement*

The ak.k12test.com and akreports.k12test.com websites are maintained in a secure and protected system, detailed in the appendix.

*Appendix 3.10 Test Site Security*

# CHAPTER 4: SCORING

## Overview

All qualified assessors complete the entire online training and proficiency testing. To become a QT, QAs participate in additional training, including administering and scoring a practice test, and reviewing an assessment and scoring procedures of a "protégé." These tools were analyzed to determine efficacy of training around scoring. The protégé tool is not included in the appendices, as this tool is used each year.

## Quality Control of Scoring

### Inter-rater Agreement

Reliability in scoring is obtained through required intensive training online, and in administering practice tests that are reviewed by a Qualified Mentor. These steps are detailed in Chapter 3.

### Handling of Exceptional Cases

Four participants in New Mentor Training required additional coaching to pass the task of assessing a protégé's test administration.

## Data Entry

After entering each student eligible for an Alaska Alternate Assessment on their caseload to the online system, assessors enter student scores into the ak.k12test.com site, on the Data Entry page.

The student's grade of enrollment preloads the possible assessments available for that student. Assessors enter the scores for each item in each eligible assessment, or indicate a reason not tested.

After entering scores in all available subject areas, assessors are prompted to submit the scores to EED. There are two ways to submit scores to EED.

1. After all scores for all required assessments have been entered, the system prompts the QA to submit the data to EED. QAs may select this option to "Submit" the data at this point; or,

2. Alternately, a QA may return to the Data Entry page and mark the record as complete by choosing the appropriate status in the Status of Data Entry drop-down box in the left-hand column.

To mark the record complete, the assessor must have entered data for each subject or given a reason why the test wasn't administered. If a subject area assessment is not administered

for a student, the assessor must choose a "Reason Not Tested" for that assessment. Scores not submitted by the close of the testing window are invalidated.

Following are the "reasons not tested" that a QA or DTC would choose to alert DRA and EED about why they are not testing a student. This information is located in the Data Entry section of the online assessment system. Students may participate in one or more Alaska Alternate Assessment content areas, and may not be eligible to participate in the AA-AAs in the other content areas tested.

**1. IEP Change** This code is selected for students who have an IEP change indicating they are no longer eligible to take the Alaska Alternate Assessment in one or more content areas, and will be taking the Standards Based Assessment (SBA) instead. This code should only be selected for the content areas in which the student is not taking the Alaska Alternate Assessment.

**2. Late Entry** This code is entered for students who enter the district from out of state or from a private school after the Alaska Alternate Assessment test window opens. In order to count for the district's participation rate, the district must administer a minimum of one assessment in reading, writing, or mathematics. This code should only be selected for the content areas in which the student is not being assessed.

**3. Suspension** The student is suspended or expelled for the entire test window. If this code is selected, it automatically applies to all content areas.

**4. Other** Any other reason must be documented in a text box that will appear when the “Other” code is selected. This code should only be selected for the content areas in which the student is not assessed. Text is limited to 50 characters, including spaces.

Because the 2010-2011 testing window was longer than in the past, "Long Term Absence" is no longer an approved reason for not testing a student in the Alaska Alternate Assessment.

#### Standard Administration With or Without Accommodations

The Alaska Alternate Assessments in reading, writing, mathematics, and science are comprised of Standard test items and Expanded Levels of Support (ELOS) test items. The standard test administration uses standardized test items, student materials, and delivery instructions. The ELOS test items offer increased support and flexibility. The ELOS items are available for students who meet the criteria that are explained below. Every year, ALL students who are eligible for the Alaska Alternate Assessment must begin with the administration of the standard test tasks and items for the student's grade level. The students may use accommodations/assistive technology during testing.

**Grade Level Assessments:** Alaska Alternate Assessments for reading, writing, and mathematics are administered in grade brands: students in grades 3 and 4 take the 3/4 tests; students in grades 5 and 6 take the 5/6 tests; students in grades 7 and 8 take the 7/8

tests; and students in grades 9 and 10 take the 9/10 test. The Alaska Alternate Assessment in science is administered in grades 4, 8, and 10. Selecting the correct grade level assessment is critical as the scores for students testing in the incorrect grade level are invalidated. For students on the non-diploma alternate assessment track, there are no tests administered after grade 10. The Alaska Alternate Assessment is the alternate assessment for both the Standards Based Assessments and the Terra Nova.

**Including Student Participation and Performance:** Students taking the Alaska Alternate Assessments (including students who take the ELOS items) are counted in their school and district for Adequate Yearly Progress (AYP) in the areas of performance and participation. Individual student scores are calculated and assigned a proficiency level. The proficiency levels are: Advanced, Proficient, Below Proficient, and Far Below Proficient. The ELOS items receive scores, but the proficiency level is Far Below Proficient as the items are non-standardized. All students receive an individual student report.

**Standard Test Administration:** The intent of administering the standard test items first is to provide an opportunity for each student to show what they know and can do in the grade level skills reflected in the standard administration of the Alaska Alternate Assessment. However, if a student is non-responsive, refuses to answer, or consistently earns zero scores (following the three-task, three-item rule described below), the standard administration should be stopped and the assessor must administer the Expanded Levels of Support (ELOS) test items. The purpose of stopping the standard test administration is to avoid having to administer the entire test to students who are not yet able to demonstrate skills at that level.

**Standard Test Administration with Accommodations:** The Alaska Alternate Assessment allows for accommodations to be utilized during test administration. Accommodations for the student are determined by the student's IEP team. The Participation Guidelines recommends that an accommodation should be used in the classroom for at least three months prior to testing. This timeline is a suggestion. It is important that the student have practice with the accommodation prior to testing; how much practice will differ by student. This amount of time allows the student to become familiar with the accommodation and ensures that the accommodation is appropriate for the student.

**Accommodations** fall into the following categories:
*** Timing/Scheduling** (e.g., extended time, frequent breaks, etc.)
*** Setting** (e.g., study carrel, student's home, separate room, etc.)
*** Presentation** (e.g., repeat directions, read aloud, large print, Braille, etc.)
* Included with Presentation is **Assistive Devices/Supports** (e.g., calculator, amplification equipment, manipulatives, etc.)
*** Response** (e.g., mark answers in book, scribe records response, point, use an assistive device, etc.)

**For the 2010-2011 assessment,** assessors were asked to define the scheduling of test administration. Assessors chose one of four options:

Timing/Scheduling Accommodations:
A. This subject administered with breaks/multiple sessions
B. This subject administered with NO breaks/one session
C. Multiple-subject administration with breaks/multiple sessions
D. Multiple-subject administration with NO breaks/one session

Choose all tests administered in one session: [checkboxes, allow one or all to be chosen]
0 Reading 0 Writing 0 Math 0 Science

"Breaks" means that the student was provided frequent breaks during testing.
"Multiple sessions" means that the test was administered over several days.

Assessors employed timing and scheduling accommodations in the following amounts:

| Choice | Reading | Writing | Math | Science |
|---|---|---|---|---|
| A | 218 | 241 | 239 | 56 |
| B | 302 | 292 | 286 | 132 |
| C | 88 | 71 | 79 | 30 |
| D | 47 | 50 | 51 | 15 |

*Appendix 4.1 Timing/Scheduling Accommodations Used*

In addition, teachers of students eligible for the Alaska Alternate Assessment are encouraged to use the Alaska Alternate Assessment practice tests throughout the school year with their students. The practice tests allow the teacher/test administrator to become more comfortable manipulating the testing materials, allow the teacher to test the efficacy of accommodations with students in testing situations, and allow the teacher to develop an understanding of student stamina and tolerance for performance testing tasks. In addition, the teacher may help the student develop test-taking strategies and become comfortable with the Alaska Alternate Assessment testing format prior to administration of the official Alaska Alternate Assessment.

There is a certain amount of flexibility for the test administrator with regard to how to present student materials. In addition to altering the materials for an allowable accommodation (e.g., increasing the text size of student materials), real-life objects may be substituted for those represented in the materials. For example, an actual glass of water may be used in lieu of the drawing of a glass of water provided in the materials, if this makes the test item more accessible to the particular student. Large Print and Braille tests are also available.

The QA may position himself in any location that is most helpful for managing the assessment materials, the student's behaviors and access to the assessment materials, and the scoring protocols. This may be side by side with the student, across the table from the student, or any position that works for the assessor and the student.

Additional resources on accommodations are available:

EED Accommodations website:
http://www.eed.state.ak.us/tls/assessment/accommodations.html

National Center on Educational Outcomes accommodations website:
http://www.cehd.umn.edu/NCEO/TopicAreas/Accommodations/Accomtopic.htm

### Analysis of Accommodations Used

DRA collected data from assessor input regarding the accommodations used in the 2010-2011 assessments. For each subject area, the total number of accommodations made is reported, as well as the total number of students receiving one or more accommodations to the Alaska Alternate Assessment.

- In **Reading**, there were 24 types of accommodations made to the test administration with types of accommodations made for student responses and 4 types of accommodations made to the test materials. A total of 126 students utilized one or more type of accommodation in Reading.
- In **Writing**, there were 20 types of accommodations made to the test administration with 8 types of accommodations made for student responses and 7 types of accommodations made to the test materials. A total of 130 students utilized one or more type of accommodation in Writing.
- In **Mathematics**, there were 18 types of accommodations made to the test administration with 9 types of accommodations made for student responses and 7 types of accommodations made to the test materials. A total of 142 students utilized one or more type of accommodation in Mathematics.
- In **Science**, there were 10 types of accommodations made to the test administration with 4 types of accommodations made for student responses and 4 types of accommodations made to the test materials. A total of 28 students utilized one or more type of accommodation in Science.

*Appendix 4.2 Accommodations Used*

### Standard Administration With or Without Accommodations AND Then Switched to the ELOS

The purpose of ELOS items is to provide access to the grade level tests for all students, even those who struggle with the standard alternate assessment test items. The focus of the ELOS is on students who have very limited or emerging systems of communication (e.g., may look at a speaker when her name is called, may indicate choice between activities, may have very early pre-skills for academic areas, etc).

In each content area the assessor must administer a minimum of three tasks and three items within each task. For each of the minimum three tasks, the student must be presented with at least three items in the task before moving on to the next task. When the student scores zeros on three consecutive items in three consecutive tasks, the assessor

should stop the assessment for that content area and must administer the required number of ELOS test items.

The three task-three item rule is operationalized as follows:

Start with Task 1 of the standard administration of the alternate assessment and proceed with successive tasks. Generally, the early tasks in each content area are easier, and tasks become progressively more difficult.

- Task 1-The assessor engages the student with the first item on a task and enters a score of zero if the student has (a) no interactive behaviors or no response, (b) actively refuses to engage in the activity, or (c) gives an incorrect answer. Next, the assessor presents the second item and enters a score of zero if the student has (a) no interactive behaviors or no response, (b) actively refuses to engage in the activity, or (c) gives an incorrect answer. Finally, the assessor moves to the third item and enters a score of zero if there is no response, the student refuses, or the student gives an incorrect answer.

- Task 2-The assessor then administers the next set of items and enters a score of zero if again there is no response, the student refuses, or the student gives an incorrect answer. When there are zeros for three consecutive items in task two, the assessor stops administering items in this task and moves to the next task.

- Task 3-Finally, the assessor administers the next set of items and enters a score of zero if again there is no response, the student refuses, or the student gives an incorrect answer. When there are zeros for three consecutive items in task three, the assessor stops administering items in this task, and the assessor stops the standard assessment in this content area. The assessor must now administer the ELOS items in this content area. ELOS items may be administered immediately to complete the assessment for this content area, or at a later time.

### ELOS Administration

**Three Task-Fifteen Item Rule for the Expanded Levels of Support (ELOS) Test Item Administration:** The ELOS test items progress from simple to more difficult tasks. Each ELOS task has five items. Using professional judgment and knowledge of the student's abilities, the assessor selects three appropriate ELOS tasks and administers all five items in each task. A student is presented with a minimum of three tasks, containing five items per task, for a total of fifteen items.

If the assessor feels that a student may already have the skill a specific task measures, the assessor may mark that task as A-Already has this skill, and select a different task to begin the assessment, continuing for a total of a minimum of three tasks - fifteen items. If after beginning a more difficult task, the student receives a score of five or less, the assessor should return to an earlier task in the sequence in order to provide the student every opportunity to show what she knows and can do.

**Scoring ELOS Tasks:** ELOS tasks are scored one through four. Scores are defined in the Levels of Independence Scoring Rubric. The additional levels of support are designed to bring the student to success. Start with the least amount of additional support (e.g., the assessor asks the question and waits for the student to respond), and introduce successively greater amounts of support, as needed by the student. Drawing the student's attention to the page by pointing in general to the answer choices is not considered a gestural support. A gestural support in ELOS is when the assessor points to the correct answer ("Which one is the math problem" -- "This one (pointing to the math problem) is the math problem. Can you point to the math problem?").

**ELOS Scores:**

1- Full physical contact to elicit student response
2- Partial physical contact to elicit student response
3- Visual, Verbal, and/or Gestural Prompts to elicit student response
4- Independent: No contact and no prompting needed to elicit student response

**Assigning other codes:**
*A-Already has this skill* - If the assessor feels that a student has the skill that a specific task measures, the assessor should mark that task as *Already has this skill*, and select a different task to begin the assessment, and continue for a total of three tasks-fifteen item minimum. *A-Already has this skill* will not be accepted as fulfilling the three task fifteen item minimum.

*I - Inappropriate/ Inaccessible* based on the nature of the student's disability - If an entire task is marked *I - Inappropriate/ Inaccessible* based on the nature of the student's disability, the assessor must document the reason this item was inappropriate or inaccessible based on the student's disability in the designated online data entry text field. The assessor must then select a more appropriate task to meet the requirements of the Three Task-Fifteen Item Minimum Rule.

*Refuses* - Student refuses to respond. Student refusal means that if a student is resisting the assessment (refuses to cooperate or respond), the assessor should attempt to administer the content area test on a different occasion. If the student continues to refuse, follow the Three Task-Fifteen Item Rule to complete testing.
When a Task or Tasks have fewer than three items, assessors were instructed to interpret the 3 X 3 rule to mean "nine consecutive zeros across a minimum of three tasks."

# CHAPTER 5: STANDARDS VALIDATION

The Alaska Alternate Assessment did not undergo a standards validation analysis this year.

*(This page left blank by design)*

# CHAPTER 6: REPORTING

## Overview

A number of tables are presented in the appendix, displaying various statistics for use in interpreting the Alaska Alternate Assessment reports.

All tables and analyses are presented for subject area results in a standardized layout format: Reading, Writing, Mathematics, and Science.

Frequency counts are used to display the number and percentage of students at various grade bands. The number and percentage of students at each score value are also displayed in the appendix. An important statistic in every table is the valid N or the number of students represented in the statistic for any given measure. Means and standard deviations are used to describe the distributions at various grade bands. These two statistics should be interpreted relative to each other; ideally, the SD is less than (even half) the amount mean, which can be interpreted as reflecting an appropriate amount of variation. When the SD is close to or greater than the mean, then the distribution is difficult to describe as there appears to be as much variation as there is centeredness. Minimums and maximums reflect the smallest and largest scores obtained on the test, respectively.

Many tables have a total that simply reflects the sum of any frequency count across all categories (e.g. grade level or score value). System missing refers to the number of students who are not in that statistical calculation (either frequency or mean).

When reliability coefficients are displayed, a value is presented that varies from a low moderate decimal (in the .30-.50 range), a moderate range (.51 to .79) or a relatively high value (in the .80 to .97 range). These values represent the degree to which two variables (e.g. forms of the test or items within the test) are related. Generally, higher is better, as the information from one measure (item or form) can be used to predict another item or form. In some cases, however, the values should not be too high (e.g., when reflecting the relations among different items in the test), because it would mean that, essentially, they are duplicating the information.

This statistic, however, is a function of the number of values (in the test) that are counted (as well as the number students behind any of these values). For example, at the total test level, many items are used to calculate the coefficient; at the strand level, sufficient items are present. However, at the task level, the number of items is so few that the values are likely to be low because there simply is not enough variation present to reflect a high coefficient.

## Reporting Student Results

Two score reports are generated for each student: an Unofficial Score Report and an Official Score Report. The **Unofficial Score Report** is generated immediately on completion and submission of student scores for all eligible alternate assessments. This report is an exact accounting of the student's performance. **Official Student Reports** are released to the District Test Coordinator in mid-May, after the AYP calculations are completed. The Official Student Report reports a student's proficiency level relative to the Extended Grade Level Expectations (ExGLE).

Chapter nine fully describes the calculations, results, and reporting methodologies for AYP.

The differences between these two sets of scores are explained in a comparison chart, available in the appendix.

*Appendix 6.1 Unofficial and Official Individual Student Report Matrix*

The appendix also lists sample documents used in reporting student results, including an Unofficial Student report, Official Student Reports in Reading, Writing, Math, and Science, and Guides to educators and parents on reading and understanding student score reports.

*Appendix 6.2a Educator Reading, Writing, Math and Science Report Examples*

*Appendix 6.2b Parent Reading, Writing, Math and Science Report Examples*

## DRA Secure Reporting Website

Official Individual Student Reports were made available to each district's District Test Coordinator on May 16, 2011. Reports are downloaded from the secure Reporting Website at akreports.k12test.com. District Test Coordinators are given a secure user identification and password at an annual training, held in February.

Reports are bundled for each district by school and then by student last name.

*Appendix 6.3 Reporting Website Manual*

# CHAPTER 7: TEST VALIDITY

## Overview

The statistical data output for the next three chapters are located in the appendix in the folder for Chapter 7. The document for each subject area contains the output regarding AYP calculations, test strand descriptive statistics, operational task descriptive statistics, task item descriptive statists and reliability statistics.

## Validity

As elaborated by Messick (1989)[1], the validity argument involves a claim with evidence evaluated to make a judgment. Three essential components of assessment systems are necessary: (a) constructs (what to measure), (b) the assessment instruments and processes (approaches to measurement), and (c) use of the test results (for specific populations). To put it simply, validation is a judgment call on the degree to which each of these components is clearly defined and adequately implemented.

Validity is a unitary concept with multifaceted processes of reasoning about a desired interpretation of test scores and subsequent uses of these test scores. In this process, we want answers for two important questions. Regardless of whether the students tested have disabilities, the questions are identical: (1) How valid is our interpretation of a student's test score? and, (2) How valid is it to use these scores in an accountability system? Validity evidence may be documented at both the item and total test levels. We use the *Standards*[2] (AERA et al., 1999) in documenting evidence on content coverage, response processes, internal structure, and relations to other variables. This document follows the essential data requirements of the federal government as needed in the peer review.[3] The critical elements highlighted in that document (with examples of acceptable evidence) include (a) academic content standards, (b) academic achievement standards, (c) a statewide assessment system, (d) validity, (e) reliability, and (f) other dimensions of technical quality.

This document addresses the latter four requirements (c-f noted above), with other documents providing essential information on the standards and statewide assessment system (see technical specifications and alignment documents for information on academic content standards and the standard setting document for information on the academic achievement standards). In addressing technical documentation, we first present content evidence, then reliability, and finally address the other three areas noted in the peer review guidance: response process, internal structures, and criterion relations.

---

[1] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.

[2] American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

[3] U. S. Department of Education (2004). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*

*Content related evidence* includes information on technical specifications and the quality of review used during the design and development of the alternate assessment. In particular, we emphasized 'universal design' in developing items and tasks that would be clear enough in their presentation and sufficiently flexible in their administration to allow ALL students access. This outcome was achieved through both the item writing and reviewing in which content experts and special educators provided feedback through the stages of test development. We also summarize outcome data as a reference for understanding subsequent validity evidence for content skills and knowledge.

## Reliability

The data file was analyzed for reliability at several levels. First, at the total test level, which is the most important because Adequate Yearly Progress (AYP) is established on the basis of this score, reliability coefficients are reported for every grade band and subject area. Second, at the strand level, coefficients are reported for every grade band and subject area. The test was designed to reflect scores at this level to ensure adequate representation across the entire range of Extended Grade Level Expectations; in the official student reports, scores for every strand are reported so that parents and teachers can follow the performance and progress of students. Third, and perhaps least important, are the scores at the task level; though we report these coefficients, they are primarily directed toward the continuous improvement of the test as EED develops new field tests and integrates them into the operational test.

In the tables for total test and strands, the reliability coefficients are reported for both the entire population (ALL students) and the students who took the complete Standard administration with students who participated in the Expanded Levels of Support (ELOS) removed (with NO ELOS). This population includes students with extremely low levels of functioning with little to no interactivity or means of communication. The reason for removing this group was to investigate the influence of missing data and its potential to spuriously inflate reliability coefficients. The first step in removing this group was to integrate the ELOS data file with the standard administration file. The second step involved splitting the file on ELOS participation and removing them so that all reliability coefficients could be recomputed at each level (total test, strand, and task). This re-analysis was done for each subject area and at all grade bands.

In general, the findings indicate that the test is very reliable for decision-making (of AYP) at the total test level. Scores were quite reliable at the strand level (with only a few strands reflecting moderate coefficients, which was primarily a function of the few number of tasks involved). Finally, as expected, scores were moderately reliable at the task level, primarily because of the few items involved. Another general (and expected) finding is that the coefficients are somewhat lower when the ELOS students scores are removed from the standard administration file although the reduction is not large, as only 9-11% of the students were administered ELOS tasks/items (see the section "Item Performance: Task Difficulty [Standard Administration, No ELOS] for summary results).

Students who participate in ELOS administration are included in the participation rate reporting for AYP; however their scores are reported as Far Below Proficient for AYP performance reporting.

## Total Test (All Students)

**Reading**: Grades 3-4 (.86 for 100 students taking 39 items), grades 5-6 (.91 for 144 students taking 35 items), grades 7-8 (.93 for 135 students taking 35 items), and grades 9-10 (.89 for 115 students taking 41 items).

**Writing**: Grades 3-4 (.81 for 99 students taking 21 items), grades 5-6 (.86 for 136 students taking 11 items), grades 7-8 (.90 for 145 students taking 16 items), and grades 9-10 (.74 for 124 students taking 17 items).

**Math**: Grades 3-4 (.90 for 103 students taking 26 items), grades 5-6 (.92 for 140 students taking 51 items), grades 7-8 (.95 for 130 students taking 66 items), and grades 9-10 (.94 for 108 students taking 65 items).

**Science**: Grade 4 (.80 for 53 students taking 24 items), grade 8 (.86 for 76 students taking 24 items), and, grade 10 (.75 for 56 students taking 24 items).

In this analysis, all items used in a grade band **test** were entered for computing the reliability of the entire test: Cronbach's Alpha based on Standardized Items.

### Reading Reliability

*Reading Grades 3 and 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.34A: Identify Signs and Symbols | .723 |
| 1.34B Identify Letter Sounds | .922 |
| 1.34C: Blend Sounds | .952 |
| 2.34A: Read Passages: Story 1, Annie Goes to a Party | .506 |
| 2.34B: Read Passages: Story 2, Jimmy Rides the Bus | .803 |

*Reading Grades 5 and 6*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.56A: Read Words | .907 |
| 1.56B: Read Sentences | .924 |
| 2.56A: Read Passages: Story 1, Jill and the Zoo | .727 |
| 2.56B: Read Passages: Story 2, Jimmy Rides the Bus | .682 |
| 2.56C: Fact/Opinion: Story 2, Jimmy Rides the Bus | .146 |

*Reading Grades 7 and 8*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.78A: Read Words of Increasing Complexity | .862 |
| 1.78B: Obtain Information | .698 |
| 1.78C: Read Sentences | .947 |
| 2.78A: Read Passages: Story 1, A Day At The Lake | .755 |
| 2.78B: Read Passages: Story 2, Going to the City | .750 |

*Reading Grades 9 and 10*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.910A: Decode Words | .926 |
| 1.910B: Identify Root Words | .614 |
| 1.910C: Follow Multi-Step Directions | .961 |
| 2.910A: Read Passages: Story 1, Mount St. Helens | .762 |
| 2.910B: Fact/Opinion: Story 1, Mount St. Helens | .382 |
| 2.910C: Read Passages: Story 2, Hannah's Homework | .833 |

**Writing Reliability**

*Writing Grades 3 and 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.34A: Copy Letters | .910 |
| 1.34B: Copy Words | .942 |
| 1.34D: Matching and Sequencing Pictures | .513 |

*Writing Grades 5 and 6*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.56A: Copy Sentences | .889 |
| 1.56C: Write Words from Dictation | .944 |
| 1.56D: Write a Sentence | .942 |

*Writing Grades 7 and 8*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.78A: Write Sentences from Dictation | .934 |
| 1.78B: Conventions of Standard English | .395 |
| 1.78C: Communicate Ideas Using Words | .901 |
| 1.78D: Write a Sentence | .938 |
| 1.78E: Revise Sentences | .723 |

*Writing Grades 9 and 10*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.910A: Conventions of Standard English | .882 |
| 1.910C: Revise Writing | .746 |

**Math Reliability**

*Math grades 3 and 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.34A Copy Numbers | .927 |
| 1.34B First and Last | .767 |
| 2.34 Same and Different | .833 |
| 3.34 Identify Shapes | .929 |

*Math grades 5 and 6*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.56A Read and Write Numbers | .796 |
| 1.56B Number Line | .714 |
| 2.56 Simple Addition | .922 |
| 3.56 Reproduce Simple Patters | .770 |
| 5.56A Shorter, Longer | .899 |
| 5.56B Identify Money | .760 |
| 6.56A Identify Shapes | .720 |
| 6.56B Same or Different Shapes | .608 |

*Math grades 7 and 8*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.78A Read and Write Numbers, Identify Place Value | .490 |
| 1.78B Identify Fractions | .153 |
| 1.78C Ordering Number Line and Pictures | .809 |
| 2.78 Double Digit Addition and Subtraction | .922 |
| 3.78B Label a Set as None or Zero | .948 |
| 3.78C Understand Symbols | .755 |
| 4.78 Read Simple Graphs | .770 |
| 5.78A Identify Units of Measurement | .750 |
| 5.78B County Money | .823 |
| 5.78C Identify Money | .899 |
| 6.78A Identify Shapes/Position | .834 |
| 6.78B Match Shapes | .846 |

*Math grades 9 and 10*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.910A Identify Place Value | .851 |
| 1.910B Identify Fractions | .651 |
| 2.910A Round Numbers | .780 |
| 2.910B Double Digit Addition/Subtraction and Single Digit Multiplication | .821 |
| 3.910B Understand Symbols | .777 |
| 4.910 Read Simple Graphs | .845 |
| 5.910A Identify Units of Measurement | .746 |
| 5.910B County Money | .753 |
| 6.910A Describe and Compare Shapes, Shapes Greater Than, Less Than, Equal To | .647 |
| 6.910B Shapes Greater Than, Less Than, Equal To | .777 |
| 6.910C Lines of Symmetry | .653 |

**Science Reliability**

*Science grade 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.4: Concepts of Physical Science | .489 |
| 2.4: Concepts of Life Science | .479 |
| 3.4: Concepts of Earth Science | .437 |

*Science grade 8*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.8: Concepts of Physical Science | .583 |
| 2.8: Concepts of Life Science | .644 |
| 3.8: Concepts of Earth Science | .576 |
| 4.8: Science and Technology | .695 |

*Science grade 10*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|---|
| 1.10: Concepts of Physical Science | .392 |
| 2.10: Concepts of Life Science | .363 |
| 3.10: Concepts of Earth Science | .421 |
| 4.10: Science and Technology | .475 |

## Item Performance

### Task Difficulty (Standard Administration, No ELOS)

*Reading Grades 3 and 4*
In the task of Identify Signs and Symbols students received a score of 50% overall. Items 1, 2, 7 and 8 were difficult, with less than 50% proficiency, and Items 3, 4, 5, 6 were medium to easy, with an average of 69% proficiency.

Identify Letter Sounds and symbols contained only one difficult item (Item 8 with 40% proficiency). Items 1 – 7, 9, and 10 were shown to be easy to medium, with 63-79% proficiency. Overall, student's received an average score of 68% for all items in this task.

The task of Blend Sounds contained items of all medium to easy difficulty level. The average proficiency rate was 62-70%. Overall proficiency for this task was 67%.

Almost all students reached proficiency on the task Identify Own Name. All items were easy, with an average score of 97% proficiency.

The task of Read Passages contained two stories with comprehension questions. Story 1: Annie Goes to a Party, received an average score of 58% overall. All items were shown to be at a medium difficulty level, with over 50% proficiency on all items. Story 2: Jimmy Rides the Bus, received an average score of 63% overall. Item 4 was difficult, with less than 50% proficiency. All other items were medium to easy, with 62-74% proficiency.

*Reading Grades 5 and 6*
In the task of Read Words students received a score of 64% overall. Items 1, 2, 3, 4, 5, and 8 were medium to easy, with an average of 58-80% proficiency. Items 6 and 7 were difficult, with less than 50% proficiency.

All Read Sentences items were in the medium to easy difficulty level. The average proficiency overall was 66%. Individual item scores ranged from 59-77%.

The task of Read Passages contained two stories with comprehension questions and a Fact and Opinion section. Story 1: Jill and the Zoo, overall received 66% proficiency. Items 6, 7, and 12 were the most difficult, with scores just above 50%. All other items received 64-97%, placing them at a medium to easy difficulty level. Story 2: Jimmy Rides the Bus, received an average score of 78% overall. All items were medium to easy, with the lowest score at 69% for Items 3 and 8. The Fact and Opinion section of story 2 contained difficult items. Items 1 and 3 received less than 50% proficiency, and Item 2 received 54% proficiency.

*Reading Grades 7 and 8*
In the task of Read Words of Increasing Complexity students received a score of 74% overall. All items were medium to easy, with scores ranging from 65-80%.

Obtain Information received an overall score of 48%. Items 2, 4, and 6 were at a medium difficulty level, with an average score of 65%. Items 1, 3, 5, and 7 were difficult, with average scores below 50%.

The task of Read Sentences received an overall score of 75%. All items were at a medium to easy difficulty level, with scores between 65-80%.

The task of Read Passages contained two stories with comprehension questions. Story 1: A Day At The Lake received an overall score of 78%. All items were medium to easy, with the lowest score 72% for item 6. Story 2: Going to the City received 64% proficiency overall. Item 9 was difficult with a score of 49%. All other items received over 50% proficiency, with Items 1 and 3 the easiest with 80% and 77% proficiency.

*Reading Grades 9 and 10*
The task of Decode Words received 76% overall proficiency. All items were medium to easy, with individual item proficiency ranging from 66-89%.

Identify Root Words contained items of all medium to easy difficulty level. Item scores ranged from 60-77%, with an overall average of 71% proficiency.

Follow Multi-Step directions contained three items, all medium to easy difficulty level. The average score for this task was 69% overall.

The task of Read Passages contained two stories with comprehension questions and a Fact and Opinion section. Story 1: Mount St. Helens received 62% proficiency overall. Items 3, 8, 9, and 10 were difficult to medium, receiving 57%, 58%, 43%, and 48% proficiency. All other items were medium to easy, receiving 69-85% proficiency. The Fact and Opinion section for Story 1 contained two items at difficult to medium, with an overall score of 52%. Story 2: Hannah's Homework received 61% proficiency overall. Items 6, 9 and 10 were difficult, with an average score of 44%. All other items were difficult to medium, with over 50% proficiency.

*Appendix 7.1 Reading Statistical Data Report*

*Writing Grades 3 and 4*
In the task of Copy Letters, students received a score of 75% overall. All items were medium to easy, with the lowest score 67% for items 9 and 11 ranging to a high score of 86% for Item 5.

All items in the Copy Words task were at an easy difficulty level. Overall, students received 75%. Individual average scores were between 70-80% for each item.

The Write Own Name task contained one item. This item was easy, with an average 80% proficiency.

In the task Matching and Sequencing Pictures, students received 88% overall. Task 1 received an overall score of 83%, and Task 2 received an overall score of 87%, making both items easy.

*Writing Grades 5 and 6*
All items in the Copy Sentences task were at a medium to easy difficulty level. There are two items in the task that received 75% proficiency overall.

The Write Words from Dictation task received an average of 59% proficiency. Items 1, 4, and 5 were difficult to medium, with students receiving 50-58%. Items 2 and 3 were medium difficulty, with over 60% proficiency.

All items in the task Write a Sentence were difficult. Overall, 32% was the average score for this task. Item 3 was the most difficult, with an average score of 30%.

*Writing Grades 7 and 8*
In the task Write Sentences from Dictation, students received a score of 50% overall. All items were difficult, with Item 3 the most difficult receiving an average score of 46%.

The Conventions of Standard English task contained three items receiving 74% proficiency overall. Item 3 was at medium difficulty, with a score of 63%. Items 1 and 2 were easy, with an average of 79% proficiency.

All items in the Communicate Ideas Using Words task were easy, receiving 77% proficiency overall. Scores ranged from 73% (task 1), to 83% (task 2).

The Write a Sentence task received 42% proficiency overall. Item 1 was difficult, with 44% proficiency. Item 2 was also difficult, with 41% proficiency.

In the task Revise a Sentence, students received 78% proficiency. All four items were easy, receiving between 72-88% proficiency.

*Writing Grades 9 and 10*
The Conventions of Standard English task received 75% proficiency overall. All items were easy except Items 6 and 7, with student scores of 66% and 68% making them medium difficulty.

Write a Story contained one item with an average score of 41% overall. This task was difficult.

In the task Revise Writing, students received 81% proficiency overall. Items ranged from medium to easy, with Item 9 receiving 67% proficiency, ranging to Item 1 receiving 95% proficiency.

*Appendix 7.2 Writing Statistical Data Report*

*Math Grades 3 and 4*
In the task Copy Numbers students received an average score of 76%. In all eight items students scored above 50%. The highest was Item 7 with 88% and the lowest was Item 1 with 67%.

In the task First and Last the average score was 69%. In all six items students scored 50% or better. The highest was Item 4, with 79% and the lowest was Item 2 with 55%.

The task Count had an average of 91% correct and was only one item.

The task Same and Different had an average of 75% correct and all eight items were 68% or above. The highest was Item 8, with 91%.

For the task Identify Shapes the average was 84% and all three items were 82% or above. The highest was Item 2, with 87%.

*Math Grades 5 and 6*
The task Read and Write Numbers had an average of 89% correct. The percentages ranged from 87-95% correct.

For the task Number Line the average was 73%. All three items had scores of 57% and above, with the highest being Item 1 with 83%.

In the only item for the task Count Objects, the average was 87%.

For the one item for the task Count the average was 90%.

The task Simple Addition had an average of 72%. The scores ranged from 63-82% for all seven items.

In the task Reproduce Simple Patterns the average was 81%. All seven items were 61% or above, with the highest being Item 6 at 90%.

The task Read Simple Graphs had an overall average of 63%. In all 10 items the averages were close to 50% or above with Items 2, 5, 8 and 9 showing the most difficulty at close to or slightly above 50%. Items 1, 3, 4, 6 and 7 all were between 85-92%.

In the task Shorter and Longer the overall average for two items was 65% and the two items had little difference in percentages.

The task Identify Money had an average of 74%. The range of percentages was from Item 2, with 65% to Item 1 with 85%.

For the task Identify Shapes the overall average was 84%. The scores were 75% or above and in all seven items and the highest was Item 1 with 97%.

Same and Different Shapes had an average of 84%. In the three items the lowest was Item 2 at 78% and the highest was Item 1 at 95%.

The final task of Identify Perimeter had an average for the one item demonstrating that it was difficult, with 38% correct.

*Math Grades 7 and 8*
The task of Read and Write Numbers had an average of 73%. Item 1 had the highest percentage of 80%.

For the task of Identify Fractions the average of two items was 69%. Item 2 had a score of 50% and Item 1 was 88%.

In the task Ordering Number Line the overall average for five items was 67%. The range of scores was from Item 5 at 55% to Item 1 at 92%.

The average for the one item in Identify Skip Patterns was 64%.

For the average of the one item in Count It was 88%.

In the task Double Digit Addition and Subtraction the overall average for seven items was 63%. The scores ranged from 57% at Item 7 to 72% at Item 2.

Reproduce and Extend Simple Patterns had an overall average for three items of 65%. Item 3 showed difficulty with a score of 36% and Item 1 was the highest with 88%.

Label a Set as None or Zero had and overall average for four items of 81%. All four items were within 81-83%.

Understand Symbols had an average of 55% for two items. The two items had scores between 53-56%.

Read Simple Graphs had an average of 65% for 10 items. Item 10 showed a lot of difficulty with a score of 20%. The rest of the items ranged from 58-96%.

Identify Units of Measurement had an average of 59%. Item 7 showed the most difficulty with 48% and Items 3 and 8 had the highest scores with 71%.

Count Money had an average for two items of 53%. Item 2 showed more difficulty with a score of 48% while Item 1 had a score of 58%.

Identify Money had an overall average of 89% for four items. The four items were between 88-91%.

Identify Shapes and Positions had an overall average for eight items of 85%. All eight items were between 81-95%.

Match Shapes had an average for three items of 95%. The three items were between 93-97%.

Identify Perimeter task showed to be difficult with an average for one item of 24%.

*Math Grades 9 and 10*
The task of Identify Place Value had an average for five items of 65%. Item 2 showed the most difficulty, with 60%, while Item 4 had the highest score of 77%.

Identify Fractions had four items for an overall average of 69%. Item 2 had the lowest score of 59% and Item 1 had the highest with 86%.

Double Digit Addition-Subtraction and Single Digit Multiplication had an average for six items of 57%. Items 5 and 6 showed the most difficulty, with scores of 42% and 44%. Items 1 and 2 had the highest scores of 68% and 75%, respectively.

Extend a Pattern and Supply Missing Elements had an average of 80% for six items. Items 5 and 6 showed the most difficulty, with scores of 69% and 65%. Items 1-3 had scores about 90%.

Understand Symbols had an overall average of 57% for 4 items. The scores ranged from Item 4 at 50%, to Item 1 at 67%.

Read Simple Graphs had an overall average for 10 items of 68%. Item 3 had the lowest score of 51%, while item 5 had the highest of 87%.

Identify Units of Measurement had an overall average of 69% for nine items. Item 2 had the lowest score of 59% and Item 8 had the highest of 79%.

Count Money had an average for three items of 64%. Item 3 showed the most difficulty with a score of 59%. Item 1 had the highest score of 69%.

Describe and Compare Shapes, Greater Than, Less Than, Equal To had an overall average of 73%. Item 5 showed a great level of difficulty, with a score of 17%. All of the other items were between 54-97%.

Shapes, Greater Than, Less Than, Equal To, had an average for three items of 91%. The three items were between scores of 89-93%.

Line of Symmetry had an average for two items of 55%. The items were between 53-57%.

Identify Perimeter was difficult, with a score of 22% for the one item.

*Appendix 7.3 Mathematics Statistical Data Report*

*Science Grade 4*
In the Grade 4 task of Physical Science students received a score of 70% overall. Items 4 and 6 were difficult, with 47% and 57% proficiency. All other items were medium to easy, with over 60% proficiency.

Grade 4 Life Science included an average score of 70% overall. Items 5 and 6 were difficult, with average scores of 47% and 57% proficiency. All other items were medium to easy, with 71-93% proficiency.

Item 5 in the Grade 4 Earth Science task was difficult, with an average score of 41%. Items 1 – 4, and 6 were easy, with scores ranging from 72-91%. The task received an average score of 72% overall.

*Science Grade 8*
In the Grade 8 Physical Science task, students received a score of 71% overall. All items received over 50% proficiency, with Item 5 being the most difficult at an average score of 53%. All other items received a score of 65% or higher.

All items in the Grade 8 Life Science task were medium to easy. The overall score for this task was 75%. Item scores ranged from 64% (Item 4) to 94% (Item 1).

Grade 8 Earth Science included an average score of 78%. Item 5 was difficult, with a score of 57%. Items 1 – 4, and 6 were medium to easy, with scores ranging from 71-94%.

Item 2 in the Grade 8 Science and Technology task was difficult, receiving a score of 54% on average. All other items were easy, receiving average scores of 85% and higher. The average score for this task was 83% overall.

*Science Grade 10*
In the Grade 10 Physical Science task, students received a score of 70% overall. Item 6 was difficult, with an average score of 55%. All other items were medium to easy, with scores ranging from 62-86%.

Grade 10 Life Science included an average score of 69% overall. Items 3, 5, and 6 were difficult, with averages ranging from 50-57%. Items 1, 2, and 4 were medium to easy, with averages ranging from 82-90%.

In the Grade 10 Earth Science task, students received a score of 61% overall. Items 1, 3, and 4 were easy, with scores ranging from 74-80%. Items 2 and 6 were very difficult, receiving less than 50% proficiency, and Item 5 was medium difficulty, with a score of 60%.

All items in Grade 10 Nature of Science – Science and Technology were medium to easy. The average score for this task was 78% overall. Item scores ranged from 65% (Item 3) to 88% (Items 5 and 6).

## Item Analysis of ELOS Administration

*ELOS Reading*
In the ELOS Word Identification Skills Task 1, on average between 40-60% of students received a score of 4 (Independent). Full Physical Prompt (1) was the next most prevalent score.

ELOS Task 2 Word Identification Skills, a score of 1 (Full Physical Prompt) was most prevalent, at 50% on average. The next most prevalent score was Partial Physical Prompt (2).

In the ELOS Task 3 Word Identification Skills, 40-50% of students received either a 1 (Full Physical Prompt), or 4 (Independent).

An average 40% of students received a 1 (Full Physical Prompt) in ELOS Task 4 Word Identification Skills, otherwise a score of 4 (Independent) was most prevalent.

ELOS Reading Task 5, Word Identification Skills had the most prevalent scores of 1 (Full Physical Prompt) and 4 (Independent) at 40-50% on average.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 6 Word Identification Skills was an average of 50%. All other scores (2, 3, 4) on average were equally prevalent.

On average, over 50% of students received a score of 1 (Full Physical Prompt) on ELOS Task 7 Forming a General Understanding. A score of 3 (Visual Prompt), was the second most prevalent score for this task.

In the ELOS Task 8 Forming a General Understanding, other than Item 1, scores of 1 (Full Physical Prompt), and 2 (Partial Physical Prompt) were the most prevalent.

The ELOS Task 9 Word Identification Skills, between 30-40% of students received a score of 1 (Full Physical Prompt), otherwise a score of 4 (Independent was most prevalent).

A score of 1 (Full Physical Prompt) was most prevalent in the ELOS Task 10 Word Identification Skills. For Items 1 and 2, an average of 30% received a score of 4 (Independent).

*ELOS Writing*
In the ELOS Task 1 Write Using a Variety of Forms, on average 50-60% of students received a score of 4 (Independent). Other than Item 1, a score of 1 (Full Physical Prompt) was the most prevalent.

Other than Item 1, an average of 50% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 2 Write Using a Variety of Forms.

Between 60-70% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 3 Write Using a Variety of Forms. Otherwise a score of 4 (Independent) was the most prevalent for this task.

In the ELOS Task 4 Write Using a Variety of Forms, on average 70% of students received a score of 1 (Full Physical Prompt).

A score of 1 (Full Physical Prompt) was the most prevalent in the ELOS Task 5 Structures and Conventions of Writing. Otherwise an average of 20-25% of students received a 3 (Visual Prompt) or 4 (Independent).

In the ELOS Task 6 Write Using a Variety of Forms, an average of 40% of students received a score of 1 (Full Physical Prompt), otherwise a score of 2 (Partial Physical Prompt) or 4 (Independent) was most prevalent.

*ELOS Mathematics*
In the ELOS Numeration Task 1, other than the first item, between 25-50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

The ELOS Task 2 Estimation and Computation, around 60% of students received a score of 1 (Full Physical Prompt). Scores of 3 (Visual Prompt), and 4 (Independent) were the next most prevalent scores in this task.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 3 Measurement was between 40-60%. A score of 4 (Independent) was the second most prevalent score.

On average, the majority of students received a 1 (Full Physical Prompt), or 4 (Independent) on the ELOS Task 4 of Geometry.

In the ELOS Measurement Task 5, between 30-60% of students received a score of 1 (Full Physical Prompt). For items 1 and 2, the next most prevalent score was 4 (Independent).

In the ELOS Measurement Task 6, other than the first task, between 35-50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

In the ELOS Measurement Task 7, other than the first task, around 50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

The ELOS Task 8 Estimation and Computation, around 50% of students received a score of 1 (Full Physical Prompt). The second most prevalent score was 4 (Independent).

On average, the majority of students received a 1 (Full Physical Prompt) in the ELOS Task 9 Statistics and Probability. Other than Item 1, students received 18-20% on Items 2, 3, and 4.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 10 Numeration was between 50-60%. All other scores (2, 3, 4) on average were equally prevalent.

*ELOS Science*
In the ELOS Task 1 Concepts of Life Science, the most prevalent scores were 1 (Full Physical Prompt) or 4 (Independent). An average of 40-60% of students received 1 or 4 for this task.

Between 50-60% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 2 Concepts of Earth Science. Otherwise a 4 (Independent) was most prevalent for this task.

Between 50-60% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 3 Concepts of Earth Science. Otherwise a score of 2 (Partial Physical Prompt) was most prevalent for this task.

In the ELOS Task 4 Concepts of Life Science, a score of 1 (Full Physical Prompt) or 4 (Independent) were most prevalent. An average of 40-60% of students received a score of 1 or 4 for this task.

An average of 30-55% of students received a score of 1 (Full Physical Prompt) or 4 (Independent) in the ELOS Task 5 Concepts of Life Science.

In the ELOS Task 6 Concepts of Life Science, a score of 1 (Full Physical Prompt) or 4 (Independent) were most prevalent. An average of 30-60% of students received a score of 1 or 4 for this task.

An average of 35-60% of students received a score of 1 (Full Physical Prompt) or 4 (Independent) for the ELOS Task 7 Concepts of Life Science.

In the ELOS Task 8 Physical Science, other than Item 5, the most prevalent scores were 1 (Full Physical Prompt) or 3 (Visual Prompt). On average 30-50% of students received a 1 and 25% of students received a 3 for this task.

A score of 1 (Full Physical Prompt) was most prevalent for the ELOS Task 9 Concepts of Earth Science, with over 50% of students receiving a score of 1.

The most prevalent score of ELOS Task 10 Concepts of Earth Science was a 1 (Full Physical Prompt). Items 3 and 4 had an average of 30-40% of students receive a score of 4 (Independent).

*Appendix 7.5 ELOS Statistical Data Report*

*Appendix 7.6 ELOS Frequency Tables*

# CHAPTER 8: DESCRIPTIVE STATISTICS

Descriptive statistics were calculated for each task, in every subject area, and in both grade bands and grade levels. The upper right header of each page refers the reader to the type of descriptive statistics displayed. For instance, "Grade Band Total Test Descriptive Statistics" refers to the descriptive statistics at the total test level for each subject, while "Writing Task Descriptive Statistics (Grade Band 3/4)" refers to the descriptive statistics for writing at the task level, in grade band 3/4. The following statistics are reported in the tables in Appendix 7 (leftmost column to rightmost column).

### Reading Strand Analysis

The Alaska Alternate Reading Assessment was grouped into four grade bands: Grades 3 and 4, Grades 5 and 6, Grades 7 and 8, and Grades 9 and 10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLE's), and were organized at the strand level.

Within the Reading Grades 3 and 4 Assessment, items were grouped into the following strands: Word Identification and Form a General Understanding. On average, students received a score of 50% in the Word Identification strand, and an average of 61% in the strand Form a General Understanding.

Within the Reading Grades 5 and 6 Assessment, items were grouped into the following strands: Word Identification, Form a General Understanding, and Analyze Content and Structure. On average, students received 60% in the Word Identification strand. Sixty-six percent was the average score in the stand Form a General Understanding, and between 65-70% in the strand Analyze Content and Structure.

Within the Reading Grades 7 and 8 Assessment, items were grouped into the following strands: Word Identification, Form a General Understanding, and Analyze Content and Structure. On average, students received a score of 63% in the Word Identification strand, and an average of 70% in the strand Form a General Understanding. Students received over 50% on average in the strand Analyze Content and Structure.

Within the Reading Grades 9 and 10 Assessment, items were grouped into the following strands: Word Identification, Form a General Understanding, and Analyze Content and Structure. In the Word Identification strand, students received an average of 73%. On average, 50% was achieved in the strand Form a General Understanding, and an average of 66% in the strand Analyze Content and Structure.

### Reading Operational Item Analysis

Each grade band assessment in Reading contained several operational items. The following summarizes the average score student's achieved on operational items in each task. In the task Identify Signs and Symbols, students received 33% correct on all operational items. Letter Sounds included an average score of 64% on operational items, and Blend Sounds 66%. Almost all students reached proficiency on the operational items in Identify Own Name at 97%. The Read Passages tasks contained several operational items across grade bands with an average score of 66%. Half of the items in both the Read Words and Read Sentences tasks contained operational items, which received an average score of 61.5%. On the operational items in the tasks of Read Words of Increasing Complexity, Decode Words, and Identify Root Words, students received scores over 70% in each task. In the task Obtain Information, a score of 46% was average for all operational items.

*Appendix 7.1 Reading Statistical Data Report*

### Writing Strand Analysis

The Alaska Alternate Writing Assessment was grouped into four grade bands: Grades 3 and 4, Grades 5 and 6, Grades 7 and 8, and Grades 9 and 10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLE's), and were organized at the strand level.

Within the Writing Grades 3 and 4 Assessment, items were grouped into the following strands: Write Using a Variety of Forms. On average, students received a score of 75% in the Write Using a Variety of Forms strand.

Within the Writing Grades 5 and 6 Assessment, items were grouped into the following strands: Write Using a Variety of Forms, and Structures and Conventions of Writing. On average, students received a score of 63.5% in the Write Using a Variety of Forms strand, and between 45-52% in the strand Structures and Conventions of Writing.

Within the Writing Grades 7 and 8 Assessment, items were grouped into the following strands: Write Using a Variety of Forms, Structures and Conventions of Writing, and Revise Writing. Students received an average score of 45% in the Write Using a Variety of Forms strand. Seventy-four percent was the average score for the Structures and Conventions of Writing strand and 40% was the average score in the strand Revise Writing.

Within the Writing Grades 9 and 10 Assessment, items were grouped into the following strands: Write Using a Variety of Forms, and Structures and Conventions of Writing. In the Write Using a Variety of Forms strand, students received an average score of 40%. Students received an average of 75% in the Structures and Conventions of Writing strand.

### Writing Operational Item Analysis

Each grade band assessment in writing contained several operational items. The following summarizes the average score student's achieved on operational items in each task. In the Copy Letters task, students received 77% on all operational items. Copy Words included an average score of 75% on operational items, and Copy Sentences 72% on average. In the Write Own Name task, student's scored an average of 72%. The tasks of Write a Sentence and Write Sentences from Dictation included average scores of 36% and 45%. In the Conventions of Standard English tasks, students received 75% on operational items. In the task Write a Story, 40% was the average score on operational items.

*Appendix 7.2 Writing Statistical Data Report*

### Math Strand Analysis

The Alaska Alternate Math Assessment was grouped into four grade bands: Grades 3 and 4, Grades 5 and 6, Grades 7 and 8, and Grades 9 and 10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLE's), and were organized at the strand level.

Within the Math Grades 3 and 4 Assessment, items were grouped into the following strands: Numeration, Functions and Relations and Geometry. On average, students received a score of 82% in the Numeration strand, an average score of 78% in the strand Functions and Relations, and an average score of 84% in the Geometry strand.

Within the Math Grades 5 and 6 Assessment, items were grouped into the following strands: Numeration, Estimation and Computation, Functions and Relations, Geometry, Statistics and Probability and Measurement. On average, students received a score of 89% in the Numeration strand, an average score of 72% in the strand Estimation and Computation, 76% for the Functions and Relations strand, 67% in the Geometry strand, 63% in the Statistics and Probability, and 70% in the Measurement strand.

Within the Math Grades 7 and 8 Assessment, items were grouped into the following strands: Numeration, Estimation and Computation, Functions and Relations, Geometry, Statistics and Probability and Measurement. On average, students received a score of 75% in the Numeration strand, an average score of 66% in the strand Estimation and Computation, 68% in the Functions and Relations strand, 61% in the Geometry strand, 65% in the Functions and Relations strand and 70% in the Measurement strand.

Within the Math Grades 9 and 10 Assessment, items were grouped into the following strands: Numeration, Estimation and Computation, Functions and Relations, Geometry and Statistics and Probability. In the Numeration strand students received an average score of 67%, in the Estimation and Computation strand students received an average score of 59%, in the Functions and Relations strand students received an average score of 71%, in the Geometry strand students received and average score of 55% and an average score of 65% in the Statistics and Probability strand.

### Math Operational Item Analysis

Each grade band assessment in Math contained several operational items. The following summarizes the average score students achieved on operational items in each task.

In the task Copy Numbers, students received 77% correct on all operational items.

For the operational task Count, students averaged between 88-91% correct and for the operational task Same and Different students averaged 78% correct.

In the tasks Identify Shapes and Read and Write Numbers, the averages were 84% and 90%.

The average for Count Objects was 87% correct and for Simple Addition the average was 72%.

For Reproduce Simple Patterns, the average was 75% correct and for Read Simple Graphs the average was 63-73%.

In the task Shorter and Longer, the average was 65% and in the task Identify Money the average was 74-89% correct.

The task Identify Shape had an average of 84% correct and the task Same or Different Shapes had an average of 78% correct.

The average correct for Identify Perimeter was 22-38%, which is the lowest correct percentage in all of the tasks.

For the tasks Ordering Number Line and Identify Sip Patterns, the average was a little over 50%.

For the tasks Double Digit Addition and Subtraction and Reproduce and Extend Simple Patterns, the averages were between 62-67% correct.

For the task Label a Set as None or Zero, the correct average was 80%. In the task Understand Symbols, the correct average was about 50-57%.

The average for Identify Units of Measurement was 60% and the average for Count Money was close to 50% and up to 60% correct.

In the task Identify Shapes and Position, the average was 87% correct, and in the task Match Shapes the average was 97% correct.

The tasks Identify Place Value and Identify Fractions had correct averages at 60%.

For Round Numbers and Double Digit Addition-Subtraction and Single Digit Multiplication the averages were 60%.

The average for Extend a Patterns and Supply Missing Elements was 82% correct and Describe and Compare Shapes, Greater Than, Less Than, Equal To was 73%.

For the Line of Symmetry Task, the average was 53% correct.

*Appendix 7.3 Mathematics Statistical Data Report*

**Science Strand Analysis**

The Alaska Alternate Science Assessment was grouped into three grade level assessments: Grade 4, Grades 8, and Grade 10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLE's), and were organized at the strand level.

Within the Science Grade 4 Assessment, items were grouped into the following strands: Physical Science, Life Science, Earth Science, and Nature of Science – Science and Technology. On average, students received a score of 62% in the Physical Science strand, and an average of 58% in the strand Life Science. The Earth Science strand received an average score of 75%, and Nature of Science – Science and Technology, 62%.

Within the Science Grade 8 Assessment, items were grouped into the following strands: Physical Science, Life Science, Earth Science and Nature of Science – Science and Technology. In the Physical Science strand, students received an average score of 62%. The Life Science strand received an average score of 77%, and the Earth Science strand, 73%. On average, students received a score of 77% in the strand Nature of Science – Science and Technology.

Within the Science Grade 10 Assessment, items were grouped into the following strands: Physical Science, Life Science, Earth Science, and Nature of Science – Science and Technology. On average, students received a score of 70% in the Physical Science strand, and 65% in the Life Science strand. In the Earth Science strand, students received a score of 72% on average. An average score of 78% was achieved in the strand Nature of Science – Science and Technology.

**Science Operational Item Analysis**

Each grade assessment in science contained several operational items. The following summarizes the average score student's achieved on operational items in each task. For the Science Assessments, the task name correlates directly to the strand name.

*Science Grade 4*
In the Grade 4 Physical Science task, students received 62% correct on all operational items. The Grade 4 Life Science task included 58% proficiency on operational items, and the Grade 4 Earth Science task included 75% proficiency on operational items. In the Grade 4 Nature of Science – Science and Technology task, operational items received 62%.

*Science Grade 8*
In the Grade 8 Physical Science task, students received 62% correct on all operational items. The Grade 8 Life Science task included 80% proficiency on operational items, and the Grade 8 Earth Science task included 73% proficiency on operational items. In the Grade 8 Nature of Science – Science and Technology task, operational items received 62% proficiency.

*Science Grade 10*
In the Grade 10 Physical Science task, students received 70% correct on all operational items. The Grade 10 Life Science task included 65% proficiency on operational items, and the Grade 10 Earth Science task included 72% proficiency on operational items. In the Grade 10 Nature of Science – Science and Technology task, operational items received 78%.

*Appendix 7. 4 Science Statistical Data Report*

# CHAPTER 9: ADEQUATE YEARLY PROGRESS

Adequate Yearly Progress (AYP) results are displayed with each attained score value presented in two different ways for depicting proficiency:

1. Four categories with 1 = Far Below, 2 = Below, 3 = Proficient, and 4 = Advanced
2. Two categories with 0 = Below (with Far Below and Below collapsed) and 1 = Above (with Proficient and Advanced collapsed)

For each table, the data present: (a) the frequency of the score value (Frequency), reflecting the number of students at that score value, (b) the percentage of students (Percent), reflecting the number of students in the grade band with a score value divided by all students taking the alternate assessment, including those with missing score values or in a different grade, (c) the percentage of students (Valid Percent), reflecting the number of students who actually had values divided by only those students with a score value in that grade band, and (d) the percentage of students with score values (Cumulative Percent), reflecting a running accumulation of percentages at/below that specific score value using only students in the grade band. The 'Frequency' and 'Valid Percent' need to be the focus of interpretations.

### Reading Adequate Yearly Progress

*Reading Grades 3 and 4*
In grades 3 and 4, 37.6% of students were proficient (a score of 3), and 20.6% of students were above proficient (a score of 4). A total of 58.2% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 41.8% of student's received a B (below proficient).

*Reading Grades 5 and 6*
In grades 5 and 6, 42.1% of students received a score of 3 (proficient), and 2.7% of students received a score of 4 (above proficient) equaling a total of 44.8% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 55.2% of student's received a B (below proficient).

*Reading Grades 7 and 8*
In grades 7 and 8, 48.9% of students were proficient (A) receiving either a score of 3 (proficient) or 4 (above proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 51.1% of student's received a B (below proficient).

*Reading Grades 9 and 10*
In grades 9 and 10, 25.3% of students received a score of 3 (proficient), and 11.3% of students received a score of 4 (above proficient) equaling a total of 36.7% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 63.3% of student's received a B (below proficient).

*Appendix 7.1 Reading Statistical Data Report*

## Writing Adequate Yearly Progress

*Writing Grades 3 and 4*
In grades 3 and 4, 26.5% of students were proficient (3), and 58.8% were above proficient (4). A total of 85.3% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 14.7% of students received a B (below proficient).

*Writing Grades 5 and 6*
In grades 5 and 6, 40.4% of students received a 3 (proficient), and 29.1% received a 4 (above proficient). On average, 69.5% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 30.5% of students were given a B (below proficient).

*Writing Grades 7 and 8*
In grades 7 and 8, 37.7% of students received a 3 (proficient), and 17.8% received a 4 (above proficient). On average, 55.5% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 44.5% of students were given a B (below proficient).

*Writing Grades 9 and 10*
In grades 9 and 10, 46.5% of students received a 3 (proficient), and 9.4% received a 4 (above proficient). On average, 55.9% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 44.1% of students were given a B (below proficient).

*Appendix 7.2 Writing Statistical Data Report*

## Math Adequate Yearly Progress

*Math Grades 3 and 4*
In grades 3 and 4, 21.3% of students were proficient (a score of 3), and 48.2% of students were above proficient (a score of 4). A total of 69.5% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 30.5% of student's received a B (below proficient).

*Math Grades 5 and 6*
In grades 5 and 6, 33.7% of students received a score of 3 (proficient), and 41.8% of students received a score of 4 (above proficient) equaling a total of 75.5% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 24.5% of student's received a B (below proficient).

*Math Grades 7 and 8*
In grades 7 and 8, 24% of students received a score of 3 (proficient), and 21.8% of students received a score of 4 (above proficient) for a total of 45.8% of students receiving an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 54.2% of student's received a B (below proficient).

*Math Grades 9 and 10*
In grades 9 and 10, 18.5% of students received a score of 3 (proficient), and 23.8% of students received a score of 4 (above proficient) equaling a total 42.3% of students receiving an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 57.6% of student's received a B (below proficient).

*Appendix 7.3 Mathematics Statistical Data Report*

**Science Adequate Yearly Progress**

*Science Grade 4*
In grade 4, 44.9% of students were proficient (a score of 3), and 17.4% of students were above proficient (a score of 4). A total of 62.3% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 37.7% of student's received a B (below proficient).

*Science Grade 8*
In grade 8, 34.1% of students were proficient (a score of 3), and 23.1% of students were above proficient (a score of 4). A total of 57.1% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 42.9% of student's received a B (below proficient).

*Science Grade 10*
In grade 10, 45.2% of students were proficient (a score of 3), and 20.5% of students were above proficient (a score of 4). A total of 65.8% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 34.2% of student's received a B (below proficient).

*Appendix 7.4 Science Statistical Data Report*

**ELOS item Analysis**

*ELOS Reading*
In the ELOS Word Identification Skills Task 1, on average between 40-60% of students received a score of 4 (Independent). Full Physical Prompt (1) was the next most prevalent score.

ELOS Task 2 Word Identification Skills, a score of 1 (Full Physical Prompt) was most prevalent at 50% on average. The next most prevalent score was Partial Physical Prompt (2).

In the ELOS Task 3 Word Identification Skills, 40-50% of students received either a 1 (Full Physical Prompt), or 4 (Independent).

An average 40% of students received a 1 (Full Physical Prompt) in ELOS Task 4 Word Identification Skills, otherwise a score of 4 (Independent) was most prevalent.

ELOS Reading Task 5, Word Identification Skills had the most prevalent scores of 1 (Full Physical Prompt) and 4 (Independent) at 40-50% on average.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 6 Word Identification Skills was an average of 50%. All other scores (2, 3, 4) on average were equally prevalent.

On average, over 50% of students received a score of 1 (Full Physical Prompt) on ELOS Task 7 Forming a General Understanding. A score of 3 (Visual Prompt), was the second most prevalent score for this task.

In the ELOS Task 8 Forming a General Understanding, other than item 1, scores of 1 (Full Physical Prompt), and 2 (Partial Physical Prompt) were the most prevalent.

The ELOS task 9 Word Identification Skills, between 30-40% of students received a score of 1 (Full Physical Prompt), otherwise a score of 4 (Independent was most prevalent).

A score of 1 (Full Physical Prompt) was most prevalent in the ELOS Task 10 Word Identification Skills. For items 1 and 2, an average of 30% received a score of 4 (Independent).

*ELOS Writing*
In the ELOS Task 1 Write Using a Variety of Forms, on average 50-60% of students received a score of 4 (Independent). Other than Item 1, a score of 1 (Full Physical Prompt) was the most prevalent.

Other than Item 1, an average of 50% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 2 Write Using a Variety of Forms.

Between 60-70% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 3 Write Using a Variety of Forms. Otherwise a score of 4 (Independent) was the most prevalent for this task.

In the ELOS Task 4 Write Using a Variety of Forms, on average 70% of students received a score of 1 (Full Physical Prompt).

A score of 1 (Full Physical Prompt) was the most prevalent in the ELOS Task 5 Structures and Conventions of Writing. Otherwise an average of 20-25% of students received a 3 (Visual Prompt) or 4 (Independent).

In the ELOS Task 6 Write Using a Variety of Forms, an average of 40% of students received a score of 1 (Full Physical Prompt), otherwise a score of 2 (Partial Physical Prompt) or 4 (Independent) was most prevalent.

*ELOS Mathematics*
In the ELOS Numeration Task 1, other than the first item, between 25-50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

The ELOS Task 2 Estimation and Computation, around 60% of students received a score of 1 (Full Physical Prompt). Scores of 3 (Visual Prompt), and 4 (Independent) were the next most prevalent scores in this task.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 3 Measurement was between 40-60%. A score of 4 (Independent) was the second most prevalent score.

On average, the majority of students received a 1 (Full Physical Prompt), or 4 (Independent) on the ELOS Task 4 of Geometry.

In the ELOS Measurement Task 5, between 30-60% of students received a score of 1 (Full Physical Prompt). For Items 1 and 2, the next most prevalent score was 4 (Independent).

In the ELOS Measurement Task 6, other than the first task, between 35-50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

In the ELOS Measurement Task 7, other than the first task, around 50% of students received a score of 1 (Full Physical Prompt), otherwise Independent (4) was the most prevalent score.

The ELOS Task 8 Estimation and Computation, around 50% of students received a score of 1 (Full Physical Prompt). The second most prevalent score was 4 (Independent).

On average, the majority of students received a 1 (Full Physical Prompt) in the ELOS Task 9 Statistics and Probability. Other than Item 1, students received 18-20% on Items 2, 3, and 4.

The percentage of students who received a 1 (Full Physical Prompt) for the ELOS Task 10 Numeration was between 50-60%. All other scores (2, 3, 4) on average were equally prevalent.

*ELOS Science*
In the ELOS Task 1 Concepts of Life Science, the most prevalent scores were 1 (Full Physical Prompt) or 4 (Independent). An average of 40-60% of students received a score of 1 or 4 for this task.

Between 50-60% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 2 Concepts of Earth Science. Otherwise a score of 4 (Independent) was most prevalent for this task.

Between 50-60% of students received a score of 1 (Full Physical Prompt) in the ELOS Task 3 Concepts of Earth Science. Otherwise a score of 2 (Partial Physical Prompt) was most prevalent for this task.

In the ELOS Task 4 Concepts of Life Science, a score of 1 (Full Physical Prompt) or 4 (Independent) were most prevalent. An average of 40-60% of students received a score of 1 or 4 for this task.

An average of 30-55% of students received a score of 1 (Full Physical Prompt) or 4 (Independent) in the ELOS Task 5 Concepts of Life Science.

In the ELOS Task 6 Concepts of Life Science, a score of 1 (Full Physical Prompt) or 4 (Independent) were most prevalent. An average of 30-60% of students received a score of 1 or 4 for this task.

An average of 35-60% of students received a score of 1 (Full Physical Prompt) or 4 (Independent) for the ELOS Task 7 Concepts of Life Science.

In the ELOS Task 8 Physical Science, other than Item 5, the most prevalent scores were 1 (Full Physical Prompt) or 3 (Visual Prompt). On average 30-50% of students received a 1 and 25% of students received a 3 for this task.

A score of 1 (Full Physical Prompt) was most prevalent for the ELOS Task 9 Concepts of Earth Science, with over 50% of students receiving a score of 1.

The most prevalent score of ELOS Task 10 Concepts of Earth Science was a 1 (Full Physical Prompt). Items 3 and 4 had an average of 30-40% of students receive a score of 4 (Independent).

*Appendix 7.5 ELOS Statistical Data Report*

*Appendix 7.6 ELOS Frequency Tables*

# CHAPTER 10: PROGRAM IMPROVEMENT

## Program Evaluation

The Alaska Alternate Assessment undergoes ongoing and multiple-level evaluation of effectiveness and reliability. In addition to a Survey of Consequential Validity, DRA and EED analyze the use of the technical components of the training and score report system, verify the effectiveness of training on scoring consistency, analyze the use and appropriateness of accommodations employed in administering the assessments, and review help desk calls for areas requiring additional training.

## Summary of Consequential Survey

Assessors are asked annually to complete a survey regarding the Alternate Assessment, their instruction and curriculum, and information about themselves.

### Training and Qualifications

Approximately 40% of the assessors who responded to the survey had between six and fifteen years' experience as educators. Nearly 30% of the respondents had more than twenty-one years' experience.

### Instructional Relevance

Seventy-five assessors believe that they do not teach differently as a result of the Alternate Assessment; 59 believe that they do.

A majority of assessors who answered the question "The Alternate Assessment is accessible to my students." as Agree or Strongly Agree (136 and 24). Additionally, 130 assessors believe that their students who participate in the Alaska Alternate Assessment are improving in their academic skills, while only 14 didn't believe that.

### Teacher Demographics and Experiences

The majority of assessors have earned a Master's degree (54%), and 70% have earned a Regular Special Education certificate, with 46% holding a Professional General Education certificate.

*Appendix 10.1a Consequential Validity Survey*

*Appendix 10.1b Consequential Validity Survey Summary Answers*

*Appendix 10.1c Consequential Validity Survey Written Responses*

## Recommendations for Future Consideration

### Technological Improvements

Throughout the 2010-2011 testing year, DRA and EED recorded issues and questions regarding the Alternate Assessment and the online training, proficiency, and score entry system. Those records were the basis for technical improvements to the system.

*Appendix 10.2 Technology Changes*

### Updates on Procedures (ELOS)

As a result of comments from Alaska assessors and feedback from EED, DRA redesigned the ELOS administration books and student materials to address the following concerns:

- Made the assessment more efficient
- Ensured that assessment items and student materials are age-appropriate
- Aligned to Early Entry Points and Extended Grade Level Expectations
- Re-bundled all subject areas into grade bands to match the levels available in the standard administration
- Ordered items within tasks with the following levels of difficulty: attention, interaction, easy, medium, hard

## Recommendations for Training

During the 2010-2011 school year, assessors participated in one or more training venues:

Face-to-face training led by DRA and EED:

- New Mentor Training
- All Mentor Training

Face-to-face training led by QTs: In-district trainings led by QTs

In addition, all assessors participated in self-paced, individual training through the ak.k12test.com website, participated in online proficiency testing, entered data into the Data Entry site (including information on accommodations used), and accessed the Help Desk for a variety of issues.

Each of these venues produced data that are useful to record and analyze to determine possible impacts on training in the 2011-2012 school year.

### Training Recommendations from EED-led trainings:

Sixteen AITs participated in the September 2010 New Mentor Training. Two events indicated a need for enhanced training: 1) Several assessors completed the protégé

protocol review task together, rather than independently; and, 2) Four assessors did not score enough points on the protégé protocol review task to earn Mentor status and required additional coaching.

> **Recommendation:** Enhance instruction over the individual testing nature of the task and increase supervision. Increase training, with examples and non-examples, of a Mentor review of protégé protocol.

**Training Recommendations from Proficiency Testing:**

The ak.k12test.com site collects data as users access every tool available. A review of the number of attempts required to pass the proficiency assessment is shown below:

| Test | ≥ 3 Resets |
|---|---|
| | |
| Admin | 4 |
| Reading | 2 |
| Writing | 18 |
| Math | 14 |
| Science | 0 |
| Refresher | 3 |

The greatest number of assessors who required three or more attempts to pass a proficiency test occurred in the subject areas of Math and Writing.

> **Recommendation:** Analyze the test items that were most frequently failed for content. Enhance face-to-face and online training around those content issues.

**Training Recommendations from Data Entry:**

The 2010-2011 Web Reports section provided a tool for QTs to monitor their own QAs as they entered their student caseload and entered and submitted the assessment scores for each student (Data Entry Status Report on the Admin Tab). The Report lists Total Students, Number Complete, and Percent Complete of scores entered in each subject area.

As the testing window drew to a close, DRA monitored the number of student tests submitted to EED. On two occasions, the Mentor was monitoring his or her QAs through the Data Entry Status Report and showed 100% of tests complete, while DRA's report showed these two assessors had not submitted the scores to DRA.

> **Recommendation:** Enhance the Data Entry Status Report to show the number of test scores that have been submitted to EED, instead of the number of tests that have been completed (test scores entered and saved, but not necessarily submitted). Provide training at All Mentor Training, through the online training site, and at the January Webinar around use of this tool.

### Training Recommendations from "Accommodations Used" data:

The 2010-2011 Score Entry page of the ak.k12test.com site included an expanded prefatory section for each student and each subject area test, requiring the assessor to record the accommodations used in testing. An analysis of the responses indicated a need for additional training.

> **Recommendation:** Provide training at All Mentor Training and online around modifications versus accommodations, and focus the need to enter very brief and specific text when describing the accommodations employed.

### Training Recommendations from Help Desk Log:

During the 2011 testing window, DRA's Help Desk operator (Sevrina Tindal) answered seventy-seven inquiries from Alaska assessors for the Alaska Alternate Assessment. Issues were organized into thirteen categories, elaborated in the table below.

Most of the Help Desk calls related to operational or procedural questions and were answered quickly. Follow-up assistance was provided when necessary.

| Issue | N |
|---|---|
| Data Entry | 13 |
| ELOS * | 4 |
| Login | 3 |
| Online Training | 4 |
| Proficiency Tests * | 11 |
| Registering on ak.k12test.com | 18 |
| Scoring * | 1 |
| Student Demographics | 3 |
| Test Administration | 9 |
| Training / Proficiency tests | 2 |
| Upgrade status | 6 |
| Writing Scoring | 1 |

The three areas with the greatest number of helpdesk calls were Data Entry (13), Proficiency Tests (11), and Registering on the ak.k12test.com site (18).

Many of the calls regarding the Proficiency Tests were to report errors in the test question or explanation of answers. Discrepancies were amended and errata posted.

Several of the Data Entry questions had to do with students whose data was entered by one assessor, but who were tested by a different assessor. In some cases, both assessors entered the student, requiring one assessor to delete the student.

Questions related to registering online included: registering new users, upgrading an account, changing emails in an account, logging in, and deleting assessors who have moved or retired.

> **Recommendation:** Enhance training at All Mentor Training and online, and provide a guidance paper to mentors regarding registration of new assessors, logging in for returning assessors, deleting retired assessors, and correcting double account issues.

*Appendix 10.3a Helpdesk Log*

*Appendix 10.3b Helpdesk Issues*